The Dissertation Committee for Mark Allan Story
certifies that this is the approved version of the following dissertation:

On Approximation Structures for Nonlinear Systems

Committee:

Irwin W. Sandberg, Supervisor

Ross Baldick

Irene Gamba

Ari Arapostathis

Joydeep Ghosh

On Approximation Structures for Nonlinear Systems

by

Mark Allan Story, B.S.; M. Eng.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2006

To Haley.

Acknowledgments

I am grateful to my supervisor, Professor Irwin Sandberg, for his help with this dissertation. It has been a great pleasure to work with someone who so enjoys the beauty of System Theory, and who has explored it so extensively. I greatly appreciate his instruction, help, and patience, and I admire his relentless curiosity. His continued supervision of my dissertation during his retirement is especially appreciated.

I also appreciate the help of those who served on my dissertation committee: Professors Ari Arapostathis, Ross Baldick, and Joydeep Ghosh, of the Department of Electrical and Computer Engineering; and Irene Gamba, of the Deparment of Mathematics. Thanks are also due to Professor Ward Cheney, for his help and advice regarding Approximation Theory.

I deeply appreciate the kindness of the donor who supported my studies by funding the Engineering Doctoral Fellowship and the Microelectronics and Computer Development Fellowship. The donor wishes to remain anonymous.

My wife, Haley, has been patient through this long process, as a true companion. My parents and brothers, and their families, have been a continual encouragement. Thanks to Kapil Dandekar, Steven Weber, Andrew McFarland, Ryan Hayes, Brian Pitts, Chip Ueltschey, and many others, for making graduate school a fun time. Thanks also to Min Chang and my other co-workers at Applied Research Laboratories for their understanding and advice, and specifically to Mike Revesz, who helped proofread some of the more difficult parts of Chapter 4.

Thanks also to all my teachers and professors for their efforts. Thanks especially to my mathematics teachers at Cumberland Regional High School, who gave me a great start: Anita Brady, Andrea Kuhar, Donna Esposito, and Cheryl White.

Finally, thanks to Melanie Gulick, Michelle Belisle, and many others who often smoothed the administrative labyrinths of graduate school.

Mark Allan Story

The University of Texas at Austin
December 2006

On Approximation Structures for Nonlinear Systems

Publication No. ______________________

Mark Allan Story, Ph.D.
The University of Texas at Austin, 2006

Supervisor: Irwin W. Sandberg

It is known that structures consisting of a finite number of linear dynamic systems followed by a memoryless nonlinear system are capable of uniformly approximating the output of a broad class of dynamic nonlinear systems arbitrarily well, over a large class of input signals. Past proofs for continuous-time systems are not constructive. In this dissertation, we show the existence of a constructive procedure for achieving such approximations for this class of systems. We give construction results for discrete-time systems as well. Also, for the first time, we give specific classes of input signals that satisfy the hypotheses of one of the more powerful prior approximation theorems. It is also shown that the members of a certain important family of feedback systems satisfy the conditions of this theorem.

Contents

List of Figures

Chapter 1

Introduction

Dynamic systems play an important role in engineering and science. The study of the theory of linear dynamic systems has generated a great many applications. For example, we know how to transmit many signals through a single channel without interference because an understanding of Fourier transform theory allows us to make use of separate orthogonal frequency bands. Using Laplace transform theory, we know how to design linear feedback amplifiers that are stable. We know how to process a continuous-time signal using discrete samples of the signal by making use of Nyquist sampling theory. Linear system theory has contributed greatly to the success of electrical engineering.

Considerably less is known about dynamic systems which are not linear. Yet a great number of problems in engineering involve dynamic nonlinear systems. For example, a need for models of nonlinear systems arises in such varied areas as semiconductor devices [1], communications hardware [2], airfoil stabilization [3], and power systems [4]. A need for nonlinear filters arises especially in image processing, in which it is necessary to remove some features of a signal, such as noise, without disturbing others, such as feature edges. Noise and feature edges are both "high-frequency" features, so it is often impossible to distinguish them using a linear filter

[5].

One approach to understanding the behavior of a nonlinear system is to "linearize" the system, i.e., to try to approximate the behavior of a nonlinear system using a linear system. This is sometimes sufficient for certain purposes. However the degree of accuracy of such an approximation is inherently limited. Simply increasing the complexity of the linear approximator does not yield an arbitrarily accurate approximation. Consequently, another approach is often needed.

One approach which does allow arbitrarily accurate approximation for a broad class of nonlinear systems involves approximating a system using a finite Volterra series, or a finite Volterra series-like expansion. A finite Volterra series is a sum of iterated integrals which may be viewed as a function-space power series. It approximates certain dynamic nonlinear systems in a manner similar to the way a polynomial approximates certain continuous real-valued functions defined on intervals of real numbers. Boyd and Chua have shown that certain nonlinear systems with "fading memory" may be uniformly approximated, to an arbitrarily high degree of accuracy (under a certain standard metric), by Volterra series-like expansions [6].[1] A broad range of applications have used the concept of a Volterra series. Finite Volterra series are used in [1] to model MOSFET transistors in the moderate inversion region, in [3] to study airfoil stabilization, in [2] to model large-signal effects in communications amplifiers, in [8] to examine the effect of radio frequency interference in op-amps, and in [4] to study pulse width modulation power systems.

Another approach, which also allows arbitrarily accurate approximation for a broad class of nonlinear systems, involves approximating a system using a structure of the form in Figure 1.1. We will refer to these structures as "L-N structures." An L-N structure can be viewed as a bank of linear dynamic systems, followed by

[1] A great deal has been written about Volterra series. In particular, much is known about the exact representation of nonlinear systems using Volterra-like series. For a good discussion, see Section 3.4 of [7].

a nonlinear memoryless element. Structures of this type, in a restricted context, were first considered by Wiener [9]. It was subsequently shown ([10], [11], [12], [13], [14]) that L-N structures are capable of uniformly approximating very broad classes of nonlinear systems, over large classes of inputs, to an arbitrarily high degree of accuracy, under certain standard metrics. Any degree of accuracy can be realized using some L-N structure, though the required number of linear maps $L_1, \ldots, L_n$, and therefore the number of inputs to N, may be large. Potential applications of L-N structures include applications of the Volterra series, such as those mentioned above.

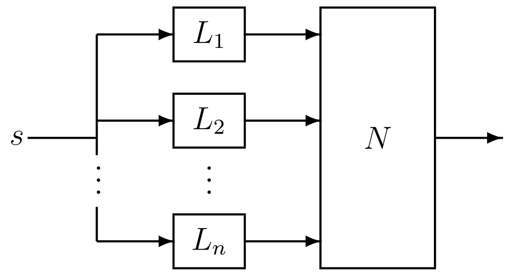

Figure 1.1: Approximating structure

A very attractive feature of the L-N structure is that its dynamic elements are separate from its nonlinear element. All of the dynamics are confined to the linear elements of the structure, and all of the nonlinearity is confined to the memoryless element. The dynamic linear elements are connected directly to the input, and the memoryless nonlinear element is connected directly to the output. This restriction of the dynamics of the L-N structure to the linear elements is particularly interesting, given the rich body of results that concern dynamics in linear systems. By contrast, every term in a finite Volterra series approximation, except the zero-order (constant) term and the first-order term, is both nonlinear and dynamic. The dynamics of the higher-order Volterra series terms are contained within iterated convolution integrals.

Another attractive feature of the L-N structure is that the nonlinear element may be constructed using any of several simple structures such as polynomials [15], sigmoidal neural networks [16], or radial basis function networks [17]. These and other approximation structures are thoroughly discussed in [18].

In this dissertation, we give several results bearing on the practical application of L-N structures. In particular, we address the following two issues in the theory of such structures. First, the theorems which guarantee the existence of L-N approximations do not tell how one can construct the approximations (except in discrete time, in the special case in which $L_1, \ldots, L_n$ are time-delay elements). It is claimed only that such an approximation exists. The proofs are based on the Stone-Weierstrass Theorem and related non-constructive results by Stone [19], and they do not indicate how one might determine the number n of linear elements, the dynamic linear maps $L_1, \ldots, L_n$, and the memoryless nonlinear map N which approximate a given dynamic nonlinear system to a given tolerance. Second, one particularly interesting result regarding L-N approximations [12] is abstractly stated in such a general way that it is not entirely clear whether there are any dynamic nonlinear systems, or any sets of inputs, that are of practical importance and that satisfy the conditions of the theorem.

Chapters 2 and 3 address the first issue described above. We give constructive approximation results in Chapter 2 for L-N structures in a discrete-time setting. It has previously been shown ([20], [21], [22], [14]) that the linear maps $L_1, \ldots, L_n$ in Figure 1.1 can be simple time-delay elements, and that the number n of time-delay elements required may be determined directly from a certain memory condition. Constructive results in this context appear in [23]. In Chapter 2, we return to the earlier idea [10] that the linear maps $L_1, \ldots, L_n$ may be more general linear filters, rather than only time-delays. We show how other combinations of linear filters may be used in place of time-delays. The number n of filters required is the same as the number of time-delay elements required in the earlier results mentioned above. However, the map N depends on the choice of $L_1, \ldots, L_n$, so it may sometimes be possible to choose $L_1, \ldots, L_n$ such that N is less complicated. We also show that n is the optimal number of filters, in a certain sense.

In Chapter 3, we give construction results for L-N structure approximations in the more challenging setting of continuous-time systems. These are the first constructive continuous-time approximation results for L-N structures. We begin by proving an approximation theorem for L-N structures, without using the non-constructive results by Stone which were the foundation of earlier approximation results ([10], [11], [12]). We then establish that the new proof is constructive by showing how one can construct an arbitrarily accurate L-N structure approximation of a certain important and well-known class of feedback systems with a uniformly Lipschitz set of input functions.[2] No additional restrictions are placed on the behavior of the system to be approximated beyond those required for existence of the approximations in earlier results.

In Chapter 4, we turn our attention to an approximation result for L-N structures which is of interest in the context of nonlinear image processing systems. Nonlinear filters are of considerable interest in image processing, largely because linear filters are unable to distinguish between typical kinds of noise, which we would like to remove, and feature edges, which we would often like to retain. (See [5] for a good overview of nonlinear image processing.) Many images may be represented by functions whose domain is in $\mathbb{R}^m$, in which m is often 2. Sometimes m is 3, as in the case of three-dimensional images [24] or moving images [25]. Typically, functions representing images have discontinuities formed by feature edges which are essential to the appearance of the image. Sandberg and Xu [12] give interesting L-N structure approximation results which hold for such functions. Specifically, it is shown

[2] By this we mean there is a single Lipschitz constant that applies to every input function in the set at every time.

that arbitrarily accurate L-N structure approximations exist for a broad class of nonlinear operators with inputs and outputs that are functions whose domain is in $\mathbb{R}^m$. The approximations are valid over sets of input functions which satisfy a certain compactness property. The compactness property does not preclude functions with discontinuities from belonging to the input function set. However, no example of a set of input functions is given in [12] which satisfies the compactness property and which includes any functions with discontinuities. This leaves the results on uncomfortably abstract grounds, in the sense that we do not know whether there are any interesting sets of discontinuous functions which satisfy the required compactness property. In Chapter 4 we introduce several interesting sets having the required compactness property. The sets we introduce are defined not in terms of their topology, but in terms more tangible in the context of images, such as the number and location of feature edges which produce discontinuities.

In Chapter 5, we show that a certain important, familiar system with inputs and outputs defined on $\mathbb{R}$ is myopic, in the sense of [12].[3] This is important because no familiar system has previously been shown to be myopic in this sense, even though [12] has established interesting approximation results for such systems. More specifically, it is known that a nonlinear feedback system satisfying the circle criterion gives rise to an input-output map with inputs and outputs defined on $\mathbb{R}_+$ (the set of nonnegative real numbers). This map satisfies a certain memory condition and a certain continuity condition ([11], [26]). In Chapter 5 we show that, in an important sense, such a system may be viewed as a myopic map with inputs and outputs extended to all of $\mathbb{R}$. We also show that a map that is myopic with respect to one weighted L_p norm is L-myopic, i.e., it is myopic with respect to every weighted L_p norm.

[3] Roughly, a system is myopic if the value of the output signal at any point x is changed little by variations in values of the input signal at points far from x, or by small variations in values of the input signal at points near x.

Chapter 2

Constructive Approximation in Discrete Time

2.1 Introduction

As we indicated in Chapter 1, it has been shown ([10], [11], [12], [13], [14]) that a broad class of systems satisfying a certain memory condition, in either discrete or continuous time, can be uniformly approximated, to as small a tolerance as desired, using the structure of Figure 1.1, where $L_1, \ldots, L_n$ are dynamic linear maps and N is a memoryless nonlinear map. N may be constructed using a simple approximation structure such as a sigmoidal neural network, a radial basis function network, or a polynomial.[1] These results are very interesting in that the dynamics of the L-N approximation structure are separated from its nonlinearities. The implications for the modeling of nonlinear systems are very exciting.

The earliest results of this kind guaranteed the existence of arbitrarily accurate approximations, but did not tell how to construct the approximations. The

[1] For a collection of interesting recent results regarding approximation of continuous memoryless nonlinear maps, see [18].

existence proofs did not indicate how one might construct an L-N approximation either, because they were based on the Stone-Weierstrass Theorem and similar non-constructive results of Stone [19]. Later, it was shown [22] that in the discrete-time case, the linear filters $L_1, \ldots, L_n$ can be simple time-delay elements. Further, the number n of time-delay elements required may be found using a certain memory condition, and the network N may be chosen to approximate a certain nonlinear map derived from G (see Sections 2 and 7 of [21]). Constructive results in this context appear in [23]. However, these results do not show whether it is possible to use more general linear filters for $L_1, \ldots, L_n$, nor do they give an idea of whether using more general filters would be advantageous in terms of the number n of filters required.

In this chapter, we return to the idea that $L_1, \ldots, L_n$ may be more general linear filters, rather than time-delay elements only. We show how to construct L-N approximations using many other sets of filters for $L_1, \ldots, L_n$. In Section 2.2, we introduce some terminology. Theorem 2.1 of Section 2.3 shows how to construct approximations of the form of Figure (1.1) using many other combinations of linear filters besides time-delay elements. The number of filters required is the same as for the case in which $L_1, \ldots, L_n$ are time-delay elements. In Section 2.4, we give Theorem 2.2, in which we show, using a counterexample, that in fact one cannot get a better result for n, in the most general case.

2.2 Preliminaries

The following definitions are used throughout this chapter. Let $\mathbb{R}$ be the set of real numbers, and let $\mathbb{R}_+$ be the set of non-negative real numbers. For an integer n, let $\mathbb{R}^n$ be the set of vectors of n real numbers. If $x \in \mathbb{R}^n$, denote the components of x

by $x_1, \ldots, x_n$, and let $\|x\|$ be defined by

$$\|x\| = \max_{j \in \{1,\ldots,n\}} |x_j|.$$

Elements of $\mathbb{R}^n$ are also understood to be real $n \times 1$ matrices (i.e. real column vectors), and any $x \in \mathbb{R}^n$ may be expressed in terms of its elements as $[x_1, \ldots, x_n]^T$, where the superscript ^{T} denotes the matrix transpose.

Let $\mathbb{Z}$ be the set of all integers, and let $\mathbb{Z}_+$ be the set of non-negative integers. Let V stand for the set of all maps $s : \mathbb{Z}_+ \to \mathbb{R}$. Denote the set of uniformly bounded $s \in V$ by ℓ_∞, and let $\|\cdot\|_\infty$ represent the norm on ℓ_∞ given by

$$\|s\|_\infty = \sup_{k \in \mathbb{Z}_+} |s(k)|, \quad s \in \ell_\infty.$$

Our input set S is the set of all $s \in \ell_\infty$ such that $\|s\|_\infty \leq b$, where b is a positive constant.

For each $\alpha \in \mathbb{Z}_+$, define the time-delay map $T_\alpha : V \to V$ by

$$(T_\alpha s)(k) = \begin{cases} s(k-\alpha), & k \geq \alpha \\ 0, & k < \alpha \end{cases}, \qquad k \in \mathbb{Z}_+,\ s \in V.$$

We also define the following associated map. For each $\alpha \in \mathbb{Z}$, let $\tilde{T}_\alpha : V \to V$ be given by

$$(\tilde{T}_\alpha s)(k) = \begin{cases} (T_\alpha s)(k), & \alpha \geq 0 \\ s(k-\alpha), & \alpha < 0 \end{cases}, \qquad k \in \mathbb{Z}_+,\ s \in V.$$

For each $\alpha, k \in \mathbb{Z}_+$, define a moving-window map $W_{k,\alpha} : V \to V$ by

$$(W_{k,\alpha}s)(j) = \begin{cases} s(j), & k-\alpha \le j \le k \\ 0, & \text{otherwise} \end{cases}, \qquad j \in \mathbb{Z}_+,\ s \in V.$$

Let G be a map from S to V. We say that G is causal if $(Gs_1)(k_0) = (Gs_2)(k_0)$ whenever $k_0 \in \mathbb{Z}_+$, $s_1, s_2 \in S$ and $s_1(k) = s_2(k)$ for every $k \le k_0$. G is said to be time-invariant if $(T_\alpha Gs)(k) = (GT_\alpha s)(k)$ whenever $k, \alpha \in \mathbb{Z}_+$ and $s \in S$. (Note that $T_\alpha s \in S$ for every $\alpha \in \mathbb{Z}_+$.) We say G has approximately finite memory if for every $\varepsilon > 0$, there is a $\Delta > 0$ such that

$$|(Gs)(k) - (GW_{k,\alpha}s)(k)| < \varepsilon \tag{2.1}$$

whenever $k \in \mathbb{Z}_+$, $\alpha \ge \Delta$, and $s \in S$.

G is said to have "continuity property $\mathcal{P}_c$" if for every $\varepsilon > 0$, there is a $\delta > 0$ such that if $s_1, s_2 \in S$ and $\|s_1 - s_2\|_\infty < \delta$, we have $|(Gs_1)(k) - (Gs_2)(k)| < \varepsilon$ for all $k \in \mathbb{Z}_+$. If G has continuity property $\mathcal{P}_c$, then the modulus of continuity of G, $\mu_G : (0,\infty) \to [0,\infty)$, is defined for each $\delta > 0$ by

$$\mu_G(\delta) = \sup\big\{|(Gs_1)(k) - (Gs_2)(k)| : s_1, s_2 \in S; \|s_1 - s_2\|_\infty < \delta;\ k \in \mathbb{Z}_+\big\}.$$

It may be difficult to find μ_G. So we say $\bar{\mu}_G : (0,\infty) \to (0,\infty)$ is a "working modulus of continuity" for G if $\bar{\mu}_G$ is a continuous, sub-additive,[2] monotonically increasing function such that $\lim_{\delta \to 0} \bar{\mu}_G(\delta) = 0$, and such that $\mu_G(\delta) \le \bar{\mu}_G(\delta)$ for every $\delta > 0$. Similarly, if $N : X \to \mathbb{R}$ is continuous, where X is a compact subset of $\mathbb{R}^n$ and where n is an integer, then the modulus of continuity of N, $\omega_N : (0,\infty) \to [0,\infty)$,

[2]By this we mean $\bar{\mu}(\delta_1 + \delta_2) \le \bar{\mu}(\delta_1) + \bar{\mu}(\delta_2)$ for all $\delta_1, \delta_2 > 0$.

is defined by

$$\omega_N(\delta) = \sup\{|Nx - Ny| : x, y \in X, \|x - y\| < \delta\}.$$

We say $\bar{\omega}_N : (0, \infty) \to [0, \infty)$ is a "working modulus of continuity" for N if $\bar{\omega}$ is a continuous, sub-additive, monotonically increasing function such that $\lim_{\delta \to 0} \bar{\omega}_N(\delta) = 0$, and such that $\omega_N(\delta) \leq \bar{\omega}_N(\delta)$ for every $\delta > 0$.

Now let $\mathcal{G}$ be the set of all $G : S \to V$ that are causal and time-invariant, and which have approximately finite memory and continuity property $\mathcal{P}_c$. For every $\varepsilon > 0$ and every positive integer Δ, let $\mathcal{G}(\varepsilon, \Delta)$ be the set of $G \in \mathcal{G}$ such that (2.1) is satisfied for all $k \in \mathbb{Z}_+$, $\alpha \geq \Delta$, and $s \in S$. Note that since every element of $\mathcal{G}$ has approximately finite memory, we have that for every $G \in \mathcal{G}$ and every $\varepsilon > 0$, there is a positive integer Δ such that $G \in \mathcal{G}(\varepsilon, \Delta)$.

We use the following maps. For each positive integer Δ, let $E_\Delta : \mathbb{R}^{\Delta+1} \to V$ and $P_\Delta : V \to \mathbb{R}^{\Delta+1}$ be given by

$$(E_\Delta x)(k) = \begin{cases} x_{(\Delta - k + 1)}, & 0 \leq k \leq \Delta \\ 0, & k > \Delta \end{cases} \qquad x \in \mathbb{R}^{\Delta+1},\ k \in \mathbb{Z}_+.$$

and

$$P_\Delta s = \left[s(\Delta), s(\Delta - 1), \ldots, s(0)\right]^T, \qquad s \in V. \tag{2.2}$$

2.3 Constructive Approximation Results

The following theorem shows that an L-N approximation may be constructed using $L_1, \ldots, L_n$ derived from the rows of any real invertible $n \times n$ matrix A, where $n = \Delta + 1$, and Δ is chosen using the approximation tolerance ε and the approximately finite memory property of G. If A is the identity matrix, $L_1, \ldots, L_n$ are time delays, as in [20], [21], and [22].

Theorem 2.1: Suppose $G \in \mathcal{G}$. Let $\varepsilon > 0$, and choose Δ such that $G \in \mathcal{G}(\varepsilon, \Delta)$. Also suppose $\bar{\mu}_G$ is a working modulus of continuity for G. Choose any real invertible $(\Delta+1) \times (\Delta+1)$ matrix A, and set

$$a = \sup_{1 \le i \le \Delta+1} \sum_{j=1}^{\Delta+1} |A_{i,j}|. \tag{2.3}$$

Let X be the set of all $x \in \mathbb{R}^{\Delta+1}$ for which $\|x\| \le ab$. Define $N : X \to \mathbb{R}$ by

$$Nx = (GE_\Delta A^{-1}x)(\Delta), \qquad x \in X. \tag{2.4}$$

Then for every $s \in S$ and every $k \in \mathbb{Z}_+$,

$$\left|(Gs)(k) - N\Big(\big[(L_0 s)(k), \ldots, (L_\Delta s)(k)\big]^T\Big)\right| < \varepsilon, \tag{2.5}$$

where for each i, $L_i : S \to \ell_\infty$ is the linear map given by

$$(L_i s)(k) = \sum_{j=0}^{\infty} h_i(j) s(k-j), \qquad k \in \mathbb{Z}_+$$

(with the understanding that $s(j) = 0$ for $j < 0$), with

$$h_i(j) = \begin{cases} A_{i+1,j+1}, & 0 \le j \le \Delta \\ 0, & j > \Delta \end{cases}.$$

Additionally, N is continuous, and if $\bar{\mu}_G$ is a working modulus of continuity for G, then $\bar{\omega}_N = \frac{1}{a'}\bar{\mu}_G$ is a working modulus of continuity for N, where

$$a' = \sup_{1 \le i \le \Delta+1} \sum_{j=1}^{\Delta+1} |(A^{-1})_{i,j}|, \tag{2.6}$$

.

Before proving the theorem, we note that because $X \subset \mathbb{R}^{\Delta+1}$ is compact and $N : X \to \mathbb{R}$ is continuous, there are a large number of simple structures which are known to be capable of uniformly approximating N as closely as desired. For example, radial basis function networks, ridge function networks (such as sigmoidal neural networks), and polynomials are all capable of approximating N. For examples of constructive proofs of such network approximation theorems, see Chapter 24 of [18] for certain ridge functions, Chapter 20 of [18] for Gaussian radial basis functions (among other classes of approximating functions), and [15] for polynomials.[3] Therefore we can find a Gaussian radial basis function network (or a polynomial, or a ridge function network) N' such that

$$|N(x) - N'(x)| < \varepsilon, \qquad x \in X. \tag{2.7}$$

An application of the triangle inequality immediately shows that Theorem 2.1 holds if N in (2.5) is replaced with N' and ε is replaced with 2ε.

If N' is a polynomial, we can bound the complexity of N' using material in [15]. Specifically, the proof of Theorem 4 of [15] (which is a constructive proof) shows that we can find a polynomial N' of order k satisfying (2.7) if

$$\varepsilon \leq c\bar{\omega}_N\left(\frac{(\Delta+1)^{\frac{3}{2}}}{k}\right) = \frac{c}{a'}\bar{\mu}_G\left(\frac{(\Delta+1)^{\frac{3}{2}}}{k}\right),$$

where c is a constant. This is important because it has been observed (See Section 2.2.1 of [22]) that if N' is a polynomial, the structure of Figure (1.1) is a doubly-finite Volterra series. Therefore we have a bound, similar to that of Theorem 2 of [22], on the complexity of a doubly-finite Volterra series required to approximate a nonlinear system satisfying the conditions of Theorem 2.1.

Also, notice that if A is the identity matrix and N is replaced with a neural

[3] Constructive approximation using Gaussian radial basis functions, and using polynomials, is discussed further in Section 3.5.

network N' satisfying (2.7), then the approximation structure is the "Time Delay Neural Network" of [21].

Proof: We begin by showing that N is continuous. Let $\zeta > 0$. Suppose $x_1, x_2 \in X$ and $\|x_1 - x_2\| < \frac{1}{a'}\bar{\mu}_G(\zeta)$. Then using (2.6), $\|A^{-1}(x_1 - x_2)\| < \bar{\mu}_G(\zeta)$. It follows that $\|E_\Delta A^{-1}x_1 - E_\Delta A^{-1}x_2\|_\infty < \bar{\mu}_G(\zeta)$, and since $\bar{\mu}_G$ is a working modulus of continuity for G,

$$|Nx_1 - Nx_2| = |(GE_\Delta A^{-1}x_1)(\Delta) - (GE_\Delta A^{-1}x_2)(\Delta)| < \zeta.$$

This shows that N is continuous, and further, that $\bar{\omega}_N = \frac{1}{a'}\bar{\mu}_G$ is a working modulus of continuity for N.

Now we turn our attention to (2.5). Fix $k \in \mathbb{Z}_+$ and $s \in S$, and let $y = [(L_0 s)(k), \ldots, (L_\Delta s)(k)]^T$. We can see from the definition of the L_i that

$$y = AP_\Delta \tilde{T}_{\Delta-k}s. \tag{2.8}$$

(Recall that P_Δ is defined in (2.2).) Because $|s(j)| \leq b$ for every j, the absolute value of each of the $\Delta + 1$ elements of the vector $P_\Delta \tilde{T}_{\Delta-k}s$ must be no greater than b. It follows from (2.3) and (2.8) that $|y_j| \leq ab$ for $j = 1, \ldots, \Delta + 1$. Therefore $y \in X$.

Combining (2.8) and (2.4), we have

$$Ny = (GE_\Delta P_\Delta \tilde{T}_{\Delta-k}s)(\Delta).$$

Note that $E_\Delta P_\Delta = W_{\Delta,\Delta}$, and that $W_{\Delta,\Delta}\tilde{T}_{\Delta-k} = \tilde{T}_{\Delta-k}W_{k,\Delta}$. Therefore

$$Ny = (G\tilde{T}_{\Delta-k}W_{k,\Delta}s)(\Delta).$$

Now if $k \leq \Delta$, time invariance gives us

$$Ny = (G\tilde{T}_{\Delta-k}W_{k,\Delta}s)(\Delta) = (GT_{\Delta-k}W_{k,\Delta}s)(\Delta) = (GW_{k,\Delta}s)(k).$$

On the other hand, if $k > \Delta$, $W_{k,\Delta} = T_{k-\Delta}\tilde{T}_{\Delta-k}W_{k,\Delta}$. Using time invariance again,

$$Ny = (G\tilde{T}_{\Delta-k}W_{k,\Delta}s)(\Delta) = (GT_{k-\Delta}\tilde{T}_{\Delta-k}W_{k,\Delta}s)(k) = (GW_{k,\Delta}s)(k).$$

Recall that we defined $y = [(L_0 s)(k), \ldots, (L_\Delta s)(k)]^T$. Therefore (2.5) follows from the approximately finite memory of G, and the proof is complete.

2.4 Lower Bounds on the Number of Required Terms

One consequence of Theorem 2.1 (as well as of Section 2 of [21]) is that we have an upper bound on the number of linear filters L_i needed for an L-N structure to approximate $G \in \mathcal{G}$. Specifically, if $\varepsilon > 0$, we need no more than $\Delta + 1$ linear filters, where Δ is chosen such that $G \in \mathcal{G}(\varepsilon, \Delta)$. Theorem 2.2 below may be seen as a partial inverse to this upper bound, in that it shows that there is an element of $\mathcal{G}(\varepsilon, \Delta)$ that cannot be approximated using an L-N structure with fewer than $\Delta + 1$ linear filters.[4]

Theorem 2.2: Let $\varepsilon > 0$, and let Δ be a positive integer. There is a $G \in \mathcal{G}(\varepsilon, \Delta)$ such that for every integer $n < \Delta$, every $N : \mathbb{R}^{n+1} \to \mathbb{R}$, and every set $\{L_0, \ldots, L_n\}$ of linear maps from V to $\mathbb{R}$, we can find an $s \in S$ and a $k \in \mathbb{Z}_+$ for which

$$\left|(Gs)(k) - N\Big(\big[(L_0 s)(k), \ldots, (L_n s)(k)\big]^T\Big)\right| \geq \varepsilon. \tag{2.9}$$

[4] Theorem 2.2 does not say that every element of $\mathcal{G}(\varepsilon, \Delta)$ requires $\Delta+1$ linear filters. For example, if $F : S \to V$ is memoryless, then clearly F is in every $\mathcal{G}(\varepsilon, \Delta)$. But F can be "approximated" exactly by the structure of Figure 1.1 using only one linear filter (the identity).

Proof: Define $G : S \to V$ by

$$(Gs)(k) = \frac{2\varepsilon}{b}\|W_{k,\Delta}s\|_\infty, \qquad s \in S,\ k \in \mathbb{Z}_+.$$

We will first show that $G \in \mathcal{G}(\varepsilon, \Delta)$. G is causal because $W_{k,\Delta}s_1 = W_{k,\Delta}s_2$ if $s_1, s_2 \in S$ and $s_1(j) = s_2(j)$ for $j \leq k$. G is time-invariant because $T_\alpha W_{k,\Delta} = W_{k-\alpha,\Delta}T_\alpha$, where α is any positive integer. It is immediately clear that G has continuity property $\mathcal{P}_c$. Finally, if $\alpha \geq \Delta$, then $W_{k,\Delta}W_{k,\alpha} = W_{k,\Delta}$ and, from the definition of G,

$$|(Gs)(k) - (GW_{k,\alpha}s)(k)| = 0, \qquad s \in S,\ k \in \mathbb{Z}_+.$$

So G has approximately finite memory, and we have shown that $G \in \mathcal{G}(\varepsilon, \Delta)$.

Define $\kappa : \mathbb{Z}_+ \to \mathbb{R}$ by

$$\kappa(i) = \begin{cases} 1, & i = 0 \\ 0, & i \neq 0 \end{cases}, \qquad i \in \mathbb{Z}_+.$$

Let A be the $(n+1) \times (\Delta+1)$ matrix whose components are

$$A_{i,j} = (L_{i-1}\kappa)(j-1), \qquad i = 1, \ldots, n+1,\ j = 1, \ldots, \Delta+1.$$

Because A has more columns than rows, there is a non-zero $u \in \mathbb{R}^{\Delta+1}$ such that $Au = [0, \ldots, 0]^T$. Let m be the index of a component of u which has the largest absolute value of any component of u, and let

$$v = \frac{b}{u_m}\, u.$$

Now $Av = [0, \ldots, 0]^T$, and

$$(L_i E_\Delta v)(\Delta) = 0, \qquad i = 0, \ldots, n.$$

Let $s_1 = E_\Delta v$, and let $s_0 \in S$ be given by $s(j) = 0$ for every $j \in \mathbb{Z}_+$. It follows that

$$N\Big(\big[(L_0 s_1)(\Delta), \ldots, (L_n s_1)(\Delta)\big]^T\Big) = N([0, \ldots, 0]^T)$$
$$N\Big(\big[(L_0 s_0)(\Delta), \ldots, (L_n s_0)(\Delta)\big]^T\Big) = N([0, \ldots, 0]^T).$$

Note that $\|W_{\Delta,\Delta} s_1\|_\infty = \sup_i |v_i| = v_m = b$. So

$$(G s_1)(\Delta) = 2\varepsilon$$
$$(G s_0)(\Delta) = 0.$$

If $N([0, \ldots, 0]^T) \leq \varepsilon$, (2.9) holds with $s = s_1$ and $k = \Delta$. On the other hand, if $N([0, \ldots, 0]^T) > \varepsilon$, and (2.9) holds with $s = s_0$ and $k = \Delta$. This proves Theorem 2.2.

Chapter 3

Constructive Approximation in Continuous Time

3.1 Introduction

In Chapter 1, we drew attention to theorems concerning the approximation of dynamic nonlinear systems using L-N structures of the form in Figure 1.1. As we indicated, it is known that the outputs of large classes of dynamic nonlinear systems may be uniformly approximated, over broad sets of inputs, using L-N structures consisting of a finite number of dynamic linear elements $L_1, \ldots, L_n$ followed by a memoryless nonlinear map N ([10], [11], [12], [13], [14]). N may be constructed using a simple approximation structure such as a sigmoidal neural network, a radial basis function network, or a polynomial. Constructive results have been given in a discrete-time setting ([20], [21], [22], [14]), with the linear elements chosen to be simple time delays. In Chapter 2 we considerably extended these constructive results by showing how to choose other filters for the linear elements.

In the more difficult continuous-time setting, no such constructive results have previously appeared. The proofs of previous theorems regarding the approximation properties of L-N structures ([10], [11], [12]) are based on the Stone-Weierstrass Theorem, and on related non-constructive results by Stone [19]. Therefore the proofs do not give any information as to how one might find an L-N structure that has the required approximation properties, if one is given a particular nonlinear system, input set, and error tolerance. A way of constructing L-N structures is needed if they are to be used to model nonlinear systems.

In this chapter, we give preliminary results on the construction of L-N structure approximations. We begin by giving an alternate proof of the existence half[1] of Theorem 2 of [11]. In this proof, the elements of the structure take a particular form. We then use this form to show how to construct L-N structure approximations in accordance with Theorem 2 of [11]. This is interesting because no previous results show how to construct these approximations in a continuous-time setting. Our results are preliminary in the sense that we do not claim that the complexity of approximation structures constructed as in this chapter is satisfactory, especially with regard to the number of radial basis function network terms or polynomial network terms used to implement N.

In Section 3.2, we give notation which is used throughout the chapter. In Section 3.3, we give Theorem 3.1, which is essentially the same as the existence portion (i.e. the "B $\Rightarrow$ A" portion) of Theorem 2 of [11]. We then prove Theorem 3.1 in a novel way. The generality of the theorem prevents the proof from being completely constructive, and we need to consider a specific case to show how to use the proof of Theorem 3.1 for construction. Therefore, in Sections 3.4 through 3.6, we show how to construct an approximation for a particular familiar class of nonlinear feedback systems, in a familiar setting. In Section 3.4, we choose a set of input functions, and also a certain set which is used to generate $L_1, \ldots, L_n$. Then we show that we

[1]Theorem 2 of [11] says that in a certain setting, arbitrarily accurate L-N structure approximations exist if and only if certain conditions are met. Construction is of interest in the case in which the conditions are met, and we wish to find the structure whose existence has been assured by the theorem. It is this half of the theorem which we prove in a new way.

can determine how many of these linear elements are needed to achieve a desired degree of approximation. In Section 3.5, we discuss previous constructive results for radial basis function network approximation and for polynomial approximation, either of which may be used to construct N. Due to results in [27], the construction of L-N structures using polynomials for N also addresses the long-standing problem of construction of uniform approximations using finite Volterra series. In Section 3.6, we show that, for a familiar class of nonlinear feedback systems, we have enough information to construct an L-N structure approximation for the system. In Section 3.7, we briefly discuss the complexity of the memoryless nonlinear network N.

3.2 Preliminary Definitions

3.2.1 Notation

The following definitions are used throughout this chapter. Let $\mathbb{R}$ and $\mathbb{R}_+$ be the set of real numbers and the set of nonnegative real numbers, respectively. For each positive integer n, let $\mathbb{R}^n$ be the set of vectors of n real numbers. For $z \in \mathbb{R}^n$, let z_i denote the i^{th} component of z. Define the norm $|\cdot|_2$ on $\mathbb{R}^n$ by

$$|z|_2 = \left(\sum_{i=1}^{n} |z_i|^2 \right)^{\frac{1}{2}}, \quad z \in \mathbb{R}^n.$$

With D being either $\mathbb{R}$, $\mathbb{R}_+$, or any other interval on $\mathbb{R}$, let $V(D)$ be the set of functions mapping D to $\mathbb{R}$. We think of the independent variable of elements of $V(D)$ as representing time. Let $L_2(D)$ be the set of Lebesgue measurable $v \in V(D)$ such that $\int_D |v(\tau)|^2 \, d\tau$ is finite.[2] Define $\|\cdot\|_2 : L_2(D) \to \mathbb{R}_+$ by $\|v\|_2 = \left(\int_D |v(\tau)|^2 \, d\tau\right)^{\frac{1}{2}}$

[2]As is customary, to avoid cumbersome notation, we do not differentiate between the set of functions $L_2(D)$ (as defined above) and the associated metric space (or normed linear space) $L_2(D)$. When discussing topological properties of $L_2(D)$, such as compactness, we mean the associated metric space, in which functions differing only on a set of Lebesgue measure zero are considered to be the same element of the metric space.

for each $v \in L_2(D)$. (Actually, we are using the notation $\|\cdot\|_2$ on many different spaces — one for each choice of D. We make clear which D we mean when it would not otherwise be clear from the context.) Let $L_\infty(D)$ be the set of Lebesgue measurable $v \in V(D)$ such that $\operatorname{ess\,sup}_{\tau\in D}|v(\tau)| < \infty$. Define $\|\cdot\|_\infty : L_\infty(D) \to \mathbb{R}_+$ by $\|v\|_\infty = \operatorname{ess\,sup}_{\tau\in D}|v(\tau)|$ for each $v \in L_\infty(D)$. Let $L_{\infty e}(D)$ be the set of Lebesgue measurable $v \in V(D)$ such that for any bounded interval $I \subseteq D$ and any $s \in L_{\infty e}(D)$, the restriction $s|_I$ of s to I is in $L_\infty(I)$.

For each $\alpha \geq 0$, define the usual time delay map $T_\alpha : V(\mathbb{R}_+) \to V(\mathbb{R}_+)$ by

$$(T_\alpha v)(t) = \begin{cases} v(t-\alpha), & t \geq \alpha \\ 0, & \text{otherwise} \end{cases}, \qquad t \in \mathbb{R}_+,\ v \in V(\mathbb{R}_+).$$

We also define the following associated map. For each $\alpha \in \mathbb{R}$, let $\tilde{T}_\alpha : V(\mathbb{R}_+) \to V(\mathbb{R}_+)$ be given by

$$(\tilde{T}_\alpha v)(t) = \begin{cases} (T_\alpha v)(t), & \alpha \geq 0 \\ v(t-\alpha), & \alpha < 0 \end{cases}, \qquad t \in \mathbb{R}_+,\ v \in V(\mathbb{R}_+).$$

We define a windowing function $W_{\tau,\beta} : V(D) \to V(D)$ for each $\tau, \beta > 0$ by

$$(W_{\tau,\beta} s)(t) = \begin{cases} s(t), & t \in [\tau-\beta, \tau] \\ 0, & \text{otherwise} \end{cases}, \qquad t \geq 0.$$

3.2.2 The General Setting

Let the approximation set S be a subset of $L_\infty(\mathbb{R}_+)$ (and therefore of $L_{\infty e}(\mathbb{R}_+)$) having the following properties:

(i) There is a positive number b such that $\|s\|_\infty \leq b$ for every $s \in S$;

(ii) S is "shift invariant," meaning that $\tilde{T}_\alpha s \in S$ for every $s \in S$ and every $\alpha \in \mathbb{R}$; and

(iii) S has "compact restrictions to intervals," by which we mean that for every closed finite interval $I \subseteq \mathbb{R}_+$, the set $\{x|_I : x \in S\}$ of restrictions of elements of S to I is compact in $L_2(I)$.

For each positive integer n, define $C(\mathbb{R}^n)$ as the set of continuous maps from $\mathbb{R}^n$ to $\mathbb{R}$. Let Υ_n be any set of maps $\upsilon : \mathbb{R}^n \to \mathbb{R}$ such that Υ_n is "dense on $C(\mathbb{R}^n)$ over compact sets"; by this we mean that for every compact $D \subset \mathbb{R}^n$, every $x \in C(D)$, and every $\varepsilon > 0$, there is an $\upsilon \in \Upsilon_n$ such that $|x(z) - \upsilon(z)| < \varepsilon$ for all $z \in D$. For example, Υ_n may consist of polynomials [15], sigmoidal neural networks [16], or radial basis function networks [17]. See [18] for an extensive discussion of this important approximation property.

Throughout this chapter, we are concerned with maps $G : L_{\infty e}(\mathbb{R}_+) \to V(\mathbb{R}_+)$. We think of G as a model of a dynamic nonlinear system, such that the input and output of the system are real functions of time drawn from $L_{\infty e}(\mathbb{R}_+)$ and $V(\mathbb{R}_+)$, respectively. We show how to approximate G uniformly over the approximation set S: that is, given $\varepsilon > 0$, we show how to find a map $\hat{G} : S \to V(\mathbb{R}_+)$ having the form of Figure 1.1 such that for every $t \in \mathbb{R}_+$ and every $s \in S$, $|(Gs)(t) - (\hat{G}s)(t)| < \varepsilon$.

We make the assumption that G has the following properties. We assume G is causal, by which we mean that for each $t \geq 0$, whenever $s_1, s_2 \in L_{\infty e}(\mathbb{R}_+)$ satisfy $s_1(\alpha) = s_2(\alpha)$ for every $\alpha \leq t$, we have $(Gs_1)(t) = (Gs_2)(t)$. We also suppose G is time invariant, meaning that for every $s \in L_{\infty e}(\mathbb{R}_+)$ and every $\alpha \geq 0$, $GT_\alpha s = T_\alpha G s$. Because G is time invariant, the zero input response of G is zero. This means that for every $t \in \mathbb{R}_+$, $(Gs_{\text{zero}})(t) = 0$, where s_{zero} is the element of $L_{\infty e}(\mathbb{R}_+)$ given by $s_{\text{zero}}(t) = 0$ for all $t \in \mathbb{R}_+$.

Furthermore, we assume $G : L_{\infty e}(\mathbb{R}_+) \to V(\mathbb{R}_+)$ has approximately finite memory, meaning that for every $\varepsilon > 0$, there is a $\Delta \geq 0$ such that

$$|(Gs)(t) - (GW_{t,\alpha}s)(t)| < \varepsilon$$

for all $t \in \mathbb{R}_+$, all $\alpha \leq \Delta$, and all $s \in L_\infty(\mathbb{R}_+)$ with $\|s\|_\infty \leq b$. The memory modulus $m_G : (0, \infty) \to [0, \infty)$ for any G having approximately finite memory is defined by

$$m_G(\Delta) = \sup\big\{|(Gs)(t) - (GW_{t,\alpha}s)(t)| : t \in \mathbb{R}_+, \alpha \leq \Delta, \|s\|_\infty \leq b\big\} \tag{3.1}$$

for each $\Delta > 0$. Because it may be difficult to find the supremum, we say any monotonically decreasing function $\bar{m}_G : (0, \infty) \to [0, \infty)$ such that $\lim_{\Delta\to\infty} \bar{m}_G(\Delta) = 0$, and such that $m_G(\Delta) \leq \bar{m}_G(\Delta)$ for every $\Delta > 0$, is a "working memory modulus" for G.

Finally, we assume that G has uniform continuity property $\mathcal{P}_c$. This means that for every $\varepsilon > 0$, there is a $\delta > 0$ such that if $s, s' \in L_2(\mathbb{R}_+)$ and $\|s - s'\|_2 < \delta$, then $|(Gs)(t) - (Gs')(t)| < \varepsilon$ for every $t \in \mathbb{R}_+$. For any G having this property, the modulus of continuity $\mu_G : (0, \infty) \to [0, \infty)$ is defined for each $\delta > 0$ by

$$\mu_G(\delta) = \sup\big\{|(Gs)(t) - (Gs')(t)| : s, s' \in L_2(\mathbb{R}_+), \|s - s'\|_2 < \delta, t \in \mathbb{R}_+\big\}. \tag{3.2}$$

Because it may be difficult to find the supremum, we say $\bar{\mu}_G : (0, \infty) \to [0, \infty)$ is a "working modulus of continuity" for G if $\bar{\mu}_G$ is a continuous, sub-additive,[3] monotonically increasing function such that $\lim_{\delta\to 0} \bar{\mu}_G(\delta) = 0$, and such that $\mu_G(\delta) \leq \bar{\mu}_G(\delta)$ for every $\delta > 0$. Similarly, if $N : X \to \mathbb{R}$ is uniformly continuous, where X is a subset of $\mathbb{R}^n$ and where n is an integer, then the modulus of continuity $\omega_N : (0, \infty) \to [0, \infty)$

[3]By this we mean $\bar{\mu}(\delta_1 + \delta_2) \leq \bar{\mu}(\delta_1) + \bar{\mu}(\delta_2)$ for all $\delta_1, \delta_2 > 0$.

is defined by

$$\omega_N(\delta) = \sup\big\{|N(x_1) - N(x_2)| : x_1, x_2 \in X, \|x_1 - x_2\| < \delta\big\}, \qquad \delta > 0.$$

We will say $\bar{\omega}_N : (0,\infty) \to [0,\infty)$ is a "working modulus of continuity" for N if $\bar{\omega}_N$ is a continuous, sub-additive, monotonically increasing function such that $\lim_{\delta\to 0} \bar{\omega}_N(\delta) = 0$, and such that $\omega_N(\delta) \le \bar{\omega}_N(\delta)$ for every $\delta > 0$.

3.3 A New Proof of an Approximation Theorem

The following theorem is essentially the same as the existence portion of Theorem 2 of [11], as we discussed earlier. However, the proof in [11] depends on Stone-Weierstrass results (specifically Theorem 1 of [19]), and does not give an idea how one might construct the approximation. We give a proof below which is more useful for construction. Comments concerning the construction of the approximation follow the proof, and in Sections 3.4 to 3.6 we give a full example.

3.3.1 Theorem and Proof

The following definitions are needed for the proof of Theorem 3.1. For each $\alpha \ge 0$, define a projection operator $P_\alpha : V(\mathbb{R}_+) \to V([0,\alpha])$ by

$$(P_\alpha v)(t) = v(t), \qquad t \in [0,\alpha],\ v \in V(\mathbb{R}). \tag{3.3}$$

Also define an extension operator $E_\alpha : V([0,\alpha]) \to V(\mathbb{R}_+)$ for each $\alpha \ge 0$ by

$$(E_\alpha v)(t) = \begin{cases} v(t) & t \in [0,\alpha] \\ 0 & t > \alpha \end{cases}, \qquad v \in V([0,\alpha]). \tag{3.4}$$

Theorem 3.1: Suppose $G : L_{\infty e}(\mathbb{R}_+) \to V(\mathbb{R}_+)$ and $S \subset L_{\infty e}(\mathbb{R}_+)$ are as given in Section 3.2.2. Then for every $\varepsilon > 0$ there exist an integer n, dynamic linear maps $L_1, ..., L_n : L_{\infty e}(\mathbb{R}_+) \to V(\mathbb{R}_+)$, and a memoryless nonlinear map $N \in \Upsilon_n$ such that

$$\left|(Gs)(t) - N\big[(L_1 s)(t), (L_2 s)(t), ..., (L_n s)(t)\big]\right| < \varepsilon, \tag{3.5}$$

for every $s \in S$ and every $t \in \mathbb{R}_+$.

Proof: Let $\varepsilon > 0$, and let $\varepsilon_m, \varepsilon_c, \varepsilon_n > 0$ with $\varepsilon_m + \varepsilon_c + \varepsilon_n = \varepsilon$. Let $\bar{m}_G$ and $\bar{\mu}_G$ be any working memory modulus and any working modulus of continuity for G, respectively.[4] Choose $\Delta > 0$ such that $\bar{m}_G(\Delta) < \varepsilon_m$, and choose $\delta > 0$ such that $\bar{\mu}_G(\delta) < \varepsilon_c$. Because G is time invariant, using the working memory modulus, we have that for every $s \in S$ and every $t \in \mathbb{R}_+$,

$$|(Gs)(t) - (GW_{\Delta,\Delta}\tilde{T}_{\Delta-t}s)(\Delta)| < \varepsilon_m. \tag{3.6}$$

Let $\mathcal{F} = \{f_1, f_2, ...\}$ be an orthonormal basis[5] for $L_2([0, \Delta])$ such that every element of $\mathcal{F}$ is in $L_\infty([0, \Delta])$. For each positive integer k, let the dynamic linear map L_k be given by

$$(L_k s)(t) = \int_0^t h_k(t - \tau)s(\tau)d\tau, \qquad t \in \mathbb{R}_+, \ s \in L_{\infty e}(\mathbb{R}_+), \tag{3.7}$$

where

$$h_k(t) = \begin{cases} f_k(\Delta - t), & t \in [0, \Delta] \\ 0, & t > \Delta \end{cases}. \tag{3.8}$$

[4] We know such $\bar{m}_G$ and $\bar{\mu}_G$ must exist because $\bar{m}_G$ may be m_G as given by (3.1), and $\bar{\mu}_G$ may be μ_G as given by (3.2).

[5] The Fourier series and Haar wavelets provide two examples of such a basis. Also note that any orthonormal basis in $L_2([0, \Delta])$ is countable (see Thm. 8.21 of [28]), so it is not improper to enumerate the elements of $\mathcal{F}$.

Notice that

$$(L_k s)(t) = \int_0^\Delta f_k(\tau)(P_\Delta \tilde{T}_{\Delta-t}s)(\tau)d\tau \tag{3.9}$$

and that $P_\Delta \tilde{T}_{\Delta-t}s \in L_2([0,\Delta])$. So using well-known results for orthonormal bases,[6] we can see that for every $s \in L_{\infty e}(\mathbb{R}_+)$ and every $t \in \mathbb{R}_+$,

$$\left(\sum_{k=1}^{\infty}[(L_k s)(t)]f_k\right)(\tau) = \big(P_\Delta \tilde{T}_{\Delta-t}s\big)(\tau) \tag{3.10}$$

for almost all[7] $\tau \in [0,\Delta]$, where the limit in the infinite sum is taken in $L_2([0,\Delta])$. We emphasize that for any fixed $t \in \mathbb{R}_+$ and $s \in S$, each $(L_k s)(t)$ is a real number, while $P_\Delta \tilde{T}_{\Delta-t}s$ and each f_k are elements of $L_2([0,\Delta])$.

Let S_Δ be the set of restrictions of elements of S to $[0,\Delta]$. Note that for any fixed $t \in \mathbb{R}_+$ and $s \in S$, $P_\Delta \tilde{T}_{\Delta-t}s \in S_\Delta$. Since S has compact restrictions to intervals, S_Δ is compact.[8] So by (3.10) and a theorem given in §28 of [29], there is a positive integer n such that for every $s \in S$ and every $t \in \mathbb{R}_+$,

$$\left\| P_\Delta \tilde{T}_{\Delta-t}s - \sum_{k=1}^{n}[(L_k s)(t)]f_k \right\|_2 < \delta, \tag{3.11}$$

where by $\|\cdot\|_2$, we mean the norm on $L_2([0,\Delta])$. Equivalently,

$$\left\| W_{\Delta,\Delta} \tilde{T}_{\Delta-t}s - \sum_{k=1}^{n}[(L_k s)(t)]E_\Delta f_k \right\|_2 < \delta, \tag{3.12}$$

where by $\|\cdot\|_2$, we now mean the norm on $L_2(\mathbb{R}_+)$. But since G has uniform continuity property $\mathcal{P}_c$,

$$\left| \big(GW_{\Delta,\Delta}\tilde{T}_{\Delta-t}s\big)(\Delta) - \left[G\left(\sum_{k=1}^{n}[(L_k s)(t)]E_\Delta f_k\right)\right](\Delta) \right| < \varepsilon_c. \tag{3.13}$$

[6] Such material may be found in, for example, sections 8.5 and 8.6 of [28].

[7] in the usual Lebesgue sense, i.e., all such τ except for a set of Lebesgue measure zero.

[8] Here we are treating elements of S_Δ which differ on a set of zero Lebesgue measure as though they are the same element of S_Δ.

Let $N_1 : \mathbb{R}^n \to \mathbb{R}$ be the memoryless nonlinear map given by

$$N_1(z) = \left[G\left(\sum_{k=1}^{n} z_k E_\Delta f_k\right)\right](\Delta), \quad z \in \mathbb{R}^n. \tag{3.14}$$

We can see that N_1 is continuous as follows. Recall that all norms on $\mathbb{R}^n$ are equivalent, so we show continuity relative to the norm $|\cdot|_2$. Let $z, z' \in \mathbb{R}^n$. Using uniform continuity property $\mathcal{P}_c$ of G,

$$|N_1(z) - N_1(z')| \le \bar{\mu}_G\left(\left\|\sum_{k=1}^{n} z_k E_\Delta f_k - \sum_{k=1}^{n} z'_k E_\Delta f_k\right\|_2\right) \tag{3.15}$$

For a moment, let us write $z_k = z'_k = 0$ for integers $k > n$, even though z and z' really only have n components. By Parseval's Identity (as in Theorem 3.3.1 of [30], for example),

$$\begin{aligned} |z - z'|_2 &= \left(\sum_{k=1}^{\infty}(z_k - z'_k)^2\right)^{\frac{1}{2}} \\ &= \left\|\sum_{k=1}^{\infty}(z_k - z'_k)E_\Delta f_k\right\|_2 = \left\|\sum_{k=1}^{n} z_k E_\Delta f_k - \sum_{k=1}^{n} z'_k E_\Delta f_k\right\|_2. \end{aligned} \tag{3.16}$$

Combining (3.15) and (3.16) gives

$$|N_1(z) - N_1(z')| \le \bar{\mu}_G(|z - z'|_2) \quad \text{for every } z, z' \in \mathbb{R}^n. \tag{3.17}$$

Therefore N_1 is uniformly continuous, and has working modulus of continuity $\bar{\omega}_{N_1} = \bar{\mu}_G$.

The f_k are orthonormal, so for each k,

$$\left(\int_0^\Delta |f_k(\tau)|^2 d\tau\right)^{\frac{1}{2}} = 1.$$

Also, for every $s \in S$, $|s(t)| \leq b$ for all $t \in \mathbb{R}_+$, so for every $t \in \mathbb{R}_+$,

$$\left(\int_0^\Delta \left|\left(P_\Delta \tilde{T}_{\Delta-t} s\right)(\tau)\right|^2 d\tau\right)^{\frac{1}{2}} \leq b\Delta^{\frac{1}{2}}.$$

So applying Holder's inequality (see Theorem 8.6 of [28], for example) to (3.9), we find that for every $s \in S$, every $t \in \mathbb{R}_+$, and every positive integer k,

$$|(L_k s)(t)| \leq b\Delta^{\frac{1}{2}}.$$

Therefore, the element of $\mathbb{R}^n$ with components given by $(L_1 s)(t), \ldots, (L_n s)(t)$ is always in the hypercube with sides of length $2b\Delta^{\frac{1}{2}}$, centered at the origin, i.e.

$$\big((L_1 s)(t), (L_2 s)(t), ..., (L_n s)(t)\big) \in [-b\Delta^{\frac{1}{2}}, b\Delta^{\frac{1}{2}}]^n, \qquad t \in \mathbb{R}_+. \tag{3.18}$$

Since $[-b\Delta^{\frac{1}{2}}, b\Delta^{\frac{1}{2}}]^n$ is a compact subset of $\mathbb{R}^n$, we may choose $N \in \Upsilon_n$ such that

$$|N_1(z) - N(z)| < \varepsilon_n, \quad z \in [-b\Delta^{\frac{1}{2}}, b\Delta^{\frac{1}{2}}]^n. \tag{3.19}$$

It follows from (3.14), (3.18), and (3.19) that for every $s \in S$ and every $t \in \mathbb{R}_+$,

$$\left|\left[G\left(\sum_{k=1}^{n}[(L_k s)(t)]E_\Delta f_k\right)\right](\Delta) - N[(L_1 s)(t), (L_2 s)(t), ..., (L_n s)(t)]\right| < \varepsilon_n. \tag{3.20}$$

Applying the triangle inequality to (3.6), (3.13), and (3.20) completes the proof.

3.3.2 Discussion of Construction Using the New Proof

The proof above tells us a great deal about a form an L-N structure approximation of G may take. The form taken by the L-N structure in the proof above is as follows. From equations (3.7) and (3.8), we see that the L_k are convolution operators whose convolution kernels are derived from an orthonormal basis for $L_2([0, \Delta])$. Also, from equations (3.14) and (3.19), we see that N is chosen to approximate the output of G at time Δ due to an input formed from a weighted sum of elements of the orthonormal basis, with the inputs to N acting as the weights. The structure of N_1 looks daunting, but because N_1 is a continuous map on a compact subset of $\mathbb{R}^n$, we may approximate N_1 using $N \in \Upsilon_n$. Therefore N may be a combination of simple elements, such as a radial basis function network or a polynomial.

Using the proof above, we can construct an L-N structure having the properties stated in Theorem 3.1, if we have the following information.

(i) For an input set S satisfying the conditions given in Section 3.2.2, and for a specific orthonormal basis $\mathcal{F}$ (which we may choose), we need to know how to find the integer n that satisfies (3.11).

(ii) For a specific family of approximation networks Υ_n, we need to know how to find $N \in \Upsilon_n$ that satisfies (3.19).

(iii) For the system G to be approximated, we need to know $\bar{\mu}_G$ and $\bar{m}_G$. Also, we need to be able to calculate the value of $(Gs)(\Delta)$ for any $s \in L_{\infty e}(\mathbb{R}_+)$.

In Section 3.4, we address (i) by showing how to find n satisfying (3.11) for a reasonable set of input functions and a familiar choice of $\mathcal{F}$. In Section 3.5, we address (ii) by discussing previously known constructive results for the case in which Υ_n is the set of radial basis function networks, and also for the case in which Υ_n is the set of polynomials in n variables. In Section 3.6, we address (iii) by showing that for a well-known class of nonlinear feedback systems, it is possible to find a working memory modulus $\bar{m}_G$ and a working modulus of continuity $\bar{\mu}_G$, and to calculate $(Gs)(\Delta)$ for any $s \in L_{\infty e}(\mathbb{R}_+)$. We therefore establish that an L-N structure approximation satisfying the conclusions of Theorem 3.1 can be constructed for a specific well-known and interesting class of dynamic nonlinear systems and an interesting set of input signals.

3.4 Determination of n for a Specific Input Set and Orthonormal Basis

In this section we establish that n satisfying (3.11) may be found for a reasonable set of input functions and a familiar choice of $\mathcal{F}$. Specifically, we let $\mathcal{F}$ be a Haar basis, and we let S be a uniformly bounded, uniformly Lipschitz (in a sense to be described in Section 3.4.1) set of elements of $L_{\infty e}(\mathbb{R}_+)$.

3.4.1 S: The Input Space

In this section we give a space $S \subset L_{\infty e}(\mathbb{R}_+)$ over which we may approximate $G : L_{\infty e}(\mathbb{R}_+) \to V(\mathbb{R}_+)$. Let $C_0(\mathbb{R}_+)$ be the set of continuous $x : \mathbb{R}_+ \to \mathbb{R}$. Now let $c > 0$ be a constant. Let our input set S be the set of all $x \in C_0(\mathbb{R}_+)$ such that $\sup_{t\in\mathbb{R}_+} |x(t)| \leq b$, and such that $|x(t_1) - x(t_2)| \leq c|t_1 - t_2|$ for every $t_1, t_2 \in \mathbb{R}_+$. So we may say that S is uniformly bounded by b, and is uniformly Lipschitz with constant c. Clearly $S \subset L_{\infty e}(\mathbb{R}_+)$, as required in Section 3.2.1. Further, it follows from the Arzela-Ascoli theorem (see Theorem 1 of §10 of [29], for example) that for every closed finite interval $I \subset \mathbb{R}_+$, the set $S_I = \{x|_I : x \in S\}$ is compact in the norm on $C_0(\mathbb{R}_+)$ given by $\sup_{t\in I} |x(t)|$. This norm gives a stronger topology on S_I than the $L_2(I)$ norm, so S_I is also compact in the $L_2(I)$ norm. Therefore S has compact restrictions to intervals, as required in Section 3.2.1.

3.4.2 $\mathcal{F}$: The Haar Orthonormal Basis for $L_2([0, \Delta])$

Let $\Delta > 0$. The Haar basis for $L_2([0, \Delta])$ is defined as follows. Let $\psi : \mathbb{R} \to \mathbb{R}$ be defined by

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{otherwise} \end{cases}, \qquad t \in \mathbb{R}.$$

For each pair of integers p, q with $p \geq 0$ and $0 \leq q \leq 2^p - 1$, let $\psi_{p,q} : [0, \Delta] \to \mathbb{R}$ be defined by

$$\psi_{p,q}(t) = \left(\frac{2^p}{\Delta}\right)^{\frac{1}{2}} \psi\left(\frac{2^p t - q}{\Delta}\right), \quad t \in [0, \Delta].$$

Let $\tilde{\psi} : [0, \Delta] \to \mathbb{R}$ be given by $\tilde{\psi}(t) = (\frac{1}{\Delta})^{\frac{1}{2}}$ for all $t \in [0, \Delta]$. It is known that all of the $\psi_{p,q}$, together with $\tilde{\psi}$, form a basis for $L_2([0, \Delta])$ (see, for example, Section 3.5 of [30]). Let $f_1 = \tilde{\psi}$, and for each integer $k \geq 2$, let $f_k = \psi_{p,q}$, where p and q are the unique pair of integers which satisfy

$$p \geq 0, \quad 0 \leq q \leq 2^p - 1, \quad \text{and} \quad k = 1 + q + 2^p. \tag{3.21}$$

Then let

$$\mathcal{F} = \{f_k\}_{k=1}^{\infty}.$$

3.4.3 Determination of n

For the specific S and $\mathcal{F}$ given above, we now show how to determine the number n of dynamic linear maps L_k which are needed to satisfy (3.11).

For every $x \in S$, and for each pair of integers p, q with $p \geq 0$ and $0 \leq q \leq 2^p - 1$, define $a_{x,p,q} \in \mathbb{R}$ by

$$a_{x,p,q} = \int_0^{\Delta} x(\tau)\psi_{p,q}(\tau)d\tau. \tag{3.22}$$

Similarly, define $\tilde{a}_x$ by

$$\tilde{a}_x = \int_0^{\Delta} x(\tau)\tilde{\psi}(\tau)d\tau. \tag{3.23}$$

For each nonnegative integer ℓ and each $x \in S$, define $x_\ell : \mathbb{R}_+ \to \mathbb{R}$ by

$$x_\ell(t) = \begin{cases} \tilde{a}_x\tilde{\psi}(t) + \sum_{p=0}^{\ell}\sum_{q=0}^{2^p-1} a_{x,p,q}\psi_{p,q}(t), & 0 \leq t \leq \Delta \\ 0 & t > \Delta \end{cases} \qquad t \in \mathbb{R}_+. \tag{3.24}$$

We now seek an upper bound for $\|x_\ell - W_{\Delta,\Delta}x\|_2$. Let $r = 2^{-(\ell+1)}\Delta$, and for each integer j, let I_j be the open interval $(jr, jr+r)$. For $t, \tau \in [0, \Delta]$, we have by Lemma 3.5.2 of [30] that if $t \in I_j$ and $\tau \in I_j$ for some integer j, then

$$\tilde{\psi}(t)\tilde{\psi}(\tau) + \sum_{p=0}^{\ell} \sum_{q=0}^{2^p-1} \psi_{p,q}(t)\psi_{p,q}(\tau)$$

is r^{-1}. Otherwise the sum above is zero. Using this fact and combining equations (3.22), (3.23), and (3.24), we have that for $0 \le j \le 2^{(\ell+1)} - 1$,

$$x_\ell(t) = \frac{1}{r}\int_{I_j} x(\tau)d\tau, \quad t \in I_j,$$

and therefore

$$|x_\ell(t) - x(t)| = \left|\frac{1}{r}\int_{I_j} \big(x(\tau) - x(t)\big)d\tau\right|, \quad t \in I_j.$$

Using the uniform Lipschitz constant,

$$|x_\ell(t) - x(t)| \le \frac{1}{r}\int_{I_j} c|\tau - t|d\tau = \frac{c}{2r}\big((t - jr)^2 + (jr + r - t)^2\big), \quad t \in I_j.$$

Now

$$\begin{aligned}
\big(\|x_\ell - W_{\Delta,\Delta}x\|_2\big)^2 &= \int_0^\Delta |x_\ell(t) - x(t)|^2 dt \\
&= \sum_{j=0}^{2^{(\ell+1)}-1} \int_{I_j} |x_\ell(t) - x(t)|^2 \, dt \\
&\le \sum_{j=0}^{2^{(\ell+1)}-1} \int_{I_j} \frac{c^2}{4r^2}\big((t - jr)^2 + (jr + r - t)^2\big)^2 dt.
\end{aligned}$$

But each integral is the same, apart from being shifted by jr. Therefore

$$\left(\|x_\ell - W_{\Delta,\Delta}x\|_2\right)^2 \leq \frac{c^2 2^{(\ell+1)}}{4r^2}\int_0^r \left(t^2 + (r-t)^2\right)^2 dt = \frac{7}{60}2^{(\ell+1)}r^3c^2,$$

and recalling $r = 2^{-(\ell+1)}\Delta$, we have

$$\|x_\ell - W_{\Delta,\Delta}x\|_2 \leq 2^{-(\ell+1)}c\Delta^{\frac{3}{2}}\sqrt{\frac{7}{60}}. \tag{3.25}$$

Now let ℓ be the smallest integer such that

$$2^\ell > \frac{c\Delta^{\frac{3}{2}}}{2\delta}\sqrt{\frac{7}{60}},$$

and let

$$n = 2^{\ell+1}.$$

Recalling the definition of f_k from Section 3.4.2, it follows from (3.9), (3.22), and (3.23) that

$$(L_1x)(\Delta) = \tilde{a}_x,$$
$$(L_kx)(\Delta) = a_{x,p,q}, \quad k = 2,\dots,n,$$

where p, q, and k are related by (3.21). Then from (3.24), we have

$$x_\ell = \sum_{k=1}^{n}[(L_kx)(\Delta)]E_\Delta f_k.$$

For each $s \in S$ and each $t \in \mathbb{R}_+$, setting $x = \tilde{T}_{\Delta-t}s$, it follows from the time invariance of the L_k and (3.25) that (3.12), and therefore (3.11), holds for this value of n.

3.5 Determination of N

In this section, we discuss results showing how to construct N from a combination of simple elements. Specifically, in 3.5.1 we discuss Gaussian radial basis function networks using results in [18], and in 3.5.2, we look at material on polynomial networks from [15].

3.5.1 Approximation of N_1 Using Gaussian Radial Basis Functions

In this section, we let Υ_n be the set of Gaussian radial basis function networks. Using ideas from the constructive proof of Theorem 1 in Chapter 24 of [18], we can find the element N of Υ_n that approximates N_1. The proof gives an estimate for N_1 using convolution, and then approximates the convolution by a Riemann sum. That Riemann sum is an element of Υ_n which approximates N_1 over $[-b\Delta^{\frac{1}{2}}, b\Delta^{\frac{1}{2}}]^n$. We note that Theorem 1 in Chapter 24 of [18] does not directly address Gaussian radial basis function networks, but we use ideas from its proof because the proof of Theorem 5 of Chapter 20 of [18] (which does guarantee the existence of approximations using Gaussian radial basis functions, as well as other classes of functions) is said to be similar, and is therefore omitted in [18].

We need the following definitions. Set $B = b\Delta^{\frac{1}{2}}$. Let n be some positive integer. Define the Gaussian function $\gamma : \mathbb{R}^n \to \mathbb{R}$ by

$$\gamma(z) = \frac{\exp(-|z|_2^2)}{(\sqrt{\pi})^n}, \quad z \in \mathbb{R}^n.$$

It is known that $\int_{\mathbb{R}^n} \gamma(z)dz = 1$, and clearly γ is bounded and continuous. Let Υ_n be the set of maps from $\mathbb{R}^n$ to $\mathbb{R}$ given for each $z \in \mathbb{R}^n$ by

$$\sum_{i=1}^{J} \xi_i \gamma(a_i z + d_i), \tag{3.26}$$

where J is a positive integer, and where $d_i \in \mathbb{R}^n$ and $\xi_i, a_i \in \mathbb{R}$ for $1 \leq i \leq J$. It is known [17] that Υ_n is dense on $C(\mathbb{R}^n)$ over compact sets.[9]

The proof of Theorem 1 of Chapter 24 of [18] uses Theorem 2 of Chapter 20 of [18] to approximate N_1 using a convolution integral. This theorem requires that the approximated function be bounded. Because N_1 is not required to be bounded, recall from (3.18) that we only need to approximate N_1 on $[-B, B]^n$. Define $N_1' : \mathbb{R}^n \to \mathbb{R}$ by

$$N_1'(z) = \begin{cases} N_1(z), & z \in [-2B, 2B]^n \\ 0, & \text{otherwise} \end{cases}, \qquad z \in \mathbb{R}^n.$$

Recall that the zero input response of G is zero. Clearly $\sum_{i=1}^n 0 E_\Delta f_k = s_{\text{zero}}$, so $N_1(0) = 0$. It follows then from (3.17) that

$$|N_1'(z)| \leq \bar{\mu}_G(|z|_2), \quad z \in [-2B, 2B]^n.$$

But $N_1'(z) = 0$ for $z \notin [-2B, 2B]^n$, and $\bar{\mu}_G$ is monotonically increasing, so

$$|N_1'(z)| \leq \bar{\mu}_G(2B\sqrt{n}), \quad z \in \mathbb{R}^n, \tag{3.27}$$

and N_1' is bounded. However, the theorem we are following also requires that the function be continuous, and N_1' is not continuous in general. But if we look at the proof,[10] we can see that it is sufficient that N_1' is continuous on a certain region in $\mathbb{R}^n$. Similarly, the second restriction in (3.30) below guarantees that N_1' only needs to be uniformly continuous on $[-2B, 2B]^n$. And indeed, using (3.17), N_1' is

[9] As we indicated, this fact also follows from Theorem 5 of Chapter 20 of [18].

[10] In the proof of Theorem 2 of Chapter 20 of [18], the supremum in the definition of the modulus of continuity is taken only over a closed ball.

uniformly continuous, and

$$|N_1'(z) - N_1'(z')| \leq \bar{\mu}_G(|z - z'|_2) \quad \text{for every } z, z' \in [-2B, 2B]^n. \tag{3.28}$$

Continuing to follow the proof of Theorem 2 of Chapter 20 of [18], we seek $a > 0$ such that for every $z \in [-B, B]^n$,

$$\left| \int_{\mathbb{R}^n} \gamma_a(\zeta) N_1'(z - \zeta) d\zeta - N_1'(z) \right| \leq \frac{1}{3}\varepsilon_n, \tag{3.29}$$

where for each $a > 0$, $\gamma_a : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\gamma_a(z) = a^n \gamma(az), \qquad z \in \mathbb{R}^n.$$

We need the following two facts about γ_a.

(i) For every $a > 0$,

$$\int_{\mathbb{R}^n} |\gamma_a(z)| dz = \int_{\mathbb{R}^n} \gamma_a(z) dz = 1.$$

(ii) For every $\sigma, a > 0$,

$$\int_{\mathbb{R}^n \setminus [-\sigma,\sigma]^n} |\gamma_a(z)| dz = \int_{\mathbb{R}^n} |\gamma_a(z)| dz - \int_{[-\sigma,\sigma]^n} |\gamma_a(z)| dz = 1 - (\mathrm{erf}(a\sigma))^n,$$

where a backslash indicates set subtraction,[11] and where $\mathrm{erf} : \mathbb{R}_+ \to \mathbb{R}$ is the well-known error function

$$\mathrm{erf}(\tau) = \frac{2}{\pi} \int_0^\tau e^{-t^2} dt, \qquad \tau \in \mathbb{R}_+.$$

[11] We mean that $\mathbb{R}^n \setminus [-\sigma, \sigma]^n$ is the set of all $z \in \mathbb{R}^n$ such that $z \notin [-\sigma, \sigma]^n$

Choose $\sigma > 0$ such that

$$\bar{\mu}_G(\sigma) < \frac{1}{6}\varepsilon_n \quad \text{and} \quad \sigma < B. \tag{3.30}$$

Then choose $a > 0$ such that

$$1 - (\operatorname{erf}(a\sigma))^n < \frac{\varepsilon_n}{12\bar{\mu}_G(2B\sqrt{n})}.$$

Using inequality (3.28), we follow the proof of Theorem 1 of Chapter 20 of [18], and see that (3.29) holds.

For each positive integer ℓ, define a partition Θ_ℓ of $[-2B, 2B]^n$ as follows. For every $\chi \in \mathbb{R}^n$ such that each component χ_i of χ is an integer between 0 and $\ell - 1$ inclusive, let

$$\theta_\chi = \left\{ z \in \mathbb{R}^n : 2B\frac{\chi_i}{\ell} \le z_i + B \le 2B\frac{\chi_i + 1}{\ell}, \quad i = 1, \ldots, n \right\}.$$

Then let Θ_ℓ consist of all the sets θ_χ.[12] The diameter of each θ_χ is

$$\sup\{|z - z'|_2 : z, z' \in \theta_\chi\} = \frac{2B}{\ell}\sqrt{n}. \tag{3.31}$$

By considering derivatives, we can see that for $y, z, z' \in \mathbb{R}^n$,

$$|\gamma(y - z) - \gamma(y - z')| \le \frac{e^{-\frac{1}{2}}\sqrt{2}}{(\sqrt{\pi})^n}|z - z'|_2,$$

so that for any $\theta \in \Theta_\ell$, any $z, z' \in \theta$, and any $y \in \mathbb{R}^n$,

$$|\gamma(y - z) - \gamma(y - z')| \le \frac{2Be^{-\frac{1}{2}}\sqrt{2n}}{\ell(\sqrt{\pi})^n}.$$

[12] This is not quite a proper partition, since the sets θ_χ share adjacent borders. Modifying the borders of the sets θ_χ to make Θ_ℓ a proper partition does not change the proof.

It follows from (3.27) that

$$\int_{[-2B,2B]^n} |N_1'(z)|dz \leq (4B)^n \bar{\mu}_G(2B\sqrt{n}).$$

For each $\theta \in \Theta_\ell$, let ζ_θ be any element of θ, and let

$$\xi_\theta = \int_\theta N_1'\left(\frac{\zeta}{a}\right)d\zeta. \tag{3.32}$$

Using quadrature as in the proof of Theorem 1 of chapter 24 of [18], we see that for every $z \in [-B,B]^n$,

$$\left|\sum_{\theta\in\Theta_\ell} \xi_\theta \gamma(az-\zeta_\theta) - \int_{\mathbb{R}^n} \gamma_a(\zeta)N_1'(z-\zeta)d\zeta\right| \leq \left(\frac{4aB}{\sqrt{\pi}}\right)^n \frac{2Be^{-\frac{1}{2}}\sqrt{2n}}{\ell}\bar{\mu}_G(2B\sqrt{n})\,. \tag{3.33}$$

Therefore if we choose ℓ such that

$$\ell > \left(\frac{4aB}{\sqrt{\pi}}\right)^n \frac{2Be^{-\frac{1}{2}}\sqrt{2n}}{\frac{1}{3}\varepsilon_n}\bar{\mu}_G(2B\sqrt{n}), \tag{3.34}$$

then using the triangle inequality on (3.29) and (3.33) gives

$$\left|\sum_{\theta\in\Theta_\ell} \xi_\theta \gamma(az-\zeta_\theta) - N_1'(z)\right| < \frac{2}{3}\varepsilon_n \tag{3.35}$$

for every $z \in [-B,B]^n$.

The approximation to N_1' in (3.35) above, to which we are led by imitating the proof of Theorem 1 in Chapter 24 of [18], is already of the form of (3.26). However, the values of ξ_θ are given by the integral in (3.32). This integral may be difficult to evaluate analytically when N_1 has the form of (3.14), because we would need an expression for (3.14) which allows the integral in (3.32) to have a closed form. Such a requirement might be difficult to meet in practice. Instead, for each $\theta \in \Theta_\ell$, let

$$\xi'_\theta = \left(\frac{2B}{\ell}\right)^n N'_1\left(\frac{\zeta_\theta}{a}\right). \tag{3.36}$$

Note that to find any ξ'_θ, we only need to by evaluate N_1 at one point. It follows from (3.14) that each ξ'_θ can be found by evaluating $(Gs)(\Delta)$ for a single s. By (3.17) and (3.31), for each $\theta \in \Theta_\ell$,

$$\begin{aligned}
|\xi'_\theta - \xi_\theta| &= \left|\left(\frac{2B}{\ell}\right)^n N'_1\left(\frac{\zeta_\theta}{a}\right) - \int_\theta N'_1\left(\frac{\zeta}{a}\right) d\zeta\right| \\
&= \left|\int_\theta N'_1\left(\frac{\zeta_\theta}{a}\right) - N'_1\left(\frac{\zeta}{a}\right) d\zeta\right| \\
&\le \int_\theta \bar{\mu}_G\left(\frac{1}{a}\big|\zeta_\theta - \zeta\big|_2\right) d\zeta \\
&\le \left(\frac{2B}{\ell}\right)^n \bar{\mu}_G\left(\frac{2B\sqrt{n}}{a\ell}\right),
\end{aligned}$$

Since Θ_ℓ has ℓ^n elements and $|\gamma(z)| \le (\frac{1}{\sqrt{\pi}})^n$ for every $z \in \mathbb{R}^n$, it follows that

$$\left|\sum_{\theta\in\Theta_\ell} \xi'_\theta \gamma(az - \zeta_\theta) - \sum_{\theta\in\Theta_\ell} \xi_\theta \gamma(az - \zeta_\theta)\right| \le \left(\frac{2B}{\sqrt{\pi}}\right)^n \bar{\mu}_G\left(\frac{2B\sqrt{n}}{a\ell}\right), \quad z \in \mathbb{R}^n. \tag{3.37}$$

Therefore if ℓ is chosen large enough such that (3.34) holds and

$$\left(\frac{2B}{\sqrt{\pi}}\right)^n \bar{\mu}_G\left(\frac{2B\sqrt{n}}{a\ell}\right) < \frac{1}{3}\varepsilon_n$$

also holds, then applying the triangle inequality to (3.29), (3.33), and (3.37) gives

$$\left|\sum_{\theta\in\Theta_\ell} \xi'_\theta \gamma(az - \zeta_\theta) - N'_1(z)\right| < \varepsilon_n, \qquad z \in [-B, B]^n. \tag{3.38}$$

We have therefore approximated $N'_1(z)$, and therefore $N_1(z)$, over $[-B, B]^n$ to within a tolerance of ε_n by an element of Υ_n. Since Θ_ℓ has ℓ^n elements, the number J of terms in (3.26) is ℓ^n.

3.5.2 Approximation of N_1 Using Polynomials

In this section we show how to find $N \in \Upsilon_n$ satisfying (3.19) if we let Υ_n be the set of polynomials in n dimensions. We will use several theorems from [15], with their constructive proofs. As an additional application, [27] shows how, if N is a polynomial network, one can transform an L-N structure into a Volterra series. Therefore the results of this section also address the long-standing problem regarding construction of uniform approximations using finite Volterra series. We do not claim that the resulting Volterra series has the best degree possible.

Using (3.17), we see that a working modulus of continuity for $N_1(B\,\cdot)$ is $\bar{\omega}_{N_1}(B\,\cdot) = \bar{\mu}_G(B\,\cdot)$.[13] Therefore by Theorem 4 of [15], for every positive integer ℓ, there is an ℓ^{th} degree polynomial υ_1 such that

$$\Big| N_1(Bx) - \upsilon_1(x) \Big| < A\bar{\mu}_G\left(\frac{Bn^{\frac{3}{2}}}{\ell}\right), \quad x \in [-1,1]^n, \tag{3.39}$$

where A is a constant that may be found by examining the proof. The proof of Theorem 4 of [15] is constructive. We may therefore choose ℓ large enough to make the right side of (3.39) smaller than any $\varepsilon_n > 0$, and then construct υ_1 that satisfies (3.39) in accordance with the proof of Theorem 4 of [15]. Then the polynomial $N \in \Upsilon_n$ given by

$$N(x) = \upsilon_1\left(\frac{x}{B}\right), \quad x \in [-1,1]^n \tag{3.40}$$

is an ℓ^{th} degree polynomial satisfying (3.19), because

$$|N_1(x) - N(x)| < A\bar{\mu}_G\left(\frac{Bn^{\frac{3}{2}}}{\ell}\right) < \varepsilon_n, \qquad x \in [-B,B]^n. \tag{3.41}$$

[13] Recall that $B = b\Delta^{\frac{1}{2}}$.

In the proof of Theorem 4 of [15], the polynomial υ_1 is constructed as follows. Let $\mathcal{S}_n$ represent the unit sphere $\{x' \in \mathbb{R}^{n+1} : |x'|_2 = 1\}$, and let $\mathcal{B}_n$ represent the unit ball $\{x \in \mathbb{R}^n : |x|_2 \leq 1\}$. Let $N_2 : \mathcal{B}_n \to \mathbb{R}$ be given by

$$N_2(x) = N_1(Bn^{\frac{1}{2}}x), \quad x \in \mathcal{B}_n. \tag{3.42}$$

A working modulus of continuity for N_2 is given by

$$\bar{\omega}_{N_2}(t) = \bar{\omega}_{N_1}(Bn^{\frac{1}{2}}t) = \bar{\mu}_G(Bn^{\frac{1}{2}}t), \qquad t \geq 0. \tag{3.43}$$

Then define $N_3 : \mathcal{S}_n \to \mathbb{R}$ by

$$N_3(x') = N_2(x), \quad x' \in \mathcal{S}_n, \tag{3.44}$$

where for each $x' \in \mathcal{S}_n$, $x \in \mathcal{B}_n$ is given by[14]

$$x_i = x'_i, \quad i = 1, \ldots, n.$$

N_3 has a working modulus of continuity given by

$$\bar{\omega}_{N_3}(t) = \bar{\omega}_{N_2}(t), \qquad t \geq 0. \tag{3.45}$$

The proof of Theorem 4 of [15] shows that if υ_2 is an ℓ^{th} degree polynomial on $\mathcal{B}_n$ satisfying

$$|N_2(x) - \upsilon_2(x)| < A\bar{\omega}_{N_2}\left(\frac{n}{\ell}\right), \quad x \in \mathcal{B}_n, \tag{3.46}$$

then the polynomial υ_1 defined by

$$\upsilon_1(x) = \upsilon_2\big(n^{-\frac{1}{2}}x\big), \quad x \in [-1,1]^n \tag{3.47}$$

[14]Since $x' \in \mathcal{S}_n$, it must be that $(x'_{n+1})^2 = 1 - \sum_{j=1}^n x_j^2$.

is of ℓ^{th} degree, and satisfies (3.39).

Further, suppose υ_3 is an ℓ^{th} degree polynomial on $\mathcal{S}_n$ satisfying[15]

$$|N_3(x') - \upsilon_3(x')| < \frac{A}{2}\bar{\omega}_{N_3}\left(\frac{n}{\ell}\right), \quad x' \in \mathcal{S}_n. \tag{3.48}$$

Let the polynomial υ_2 be defined by

$$\upsilon_2(x) = \tfrac{1}{2}\upsilon_3(x'^{+}) + \tfrac{1}{2}\upsilon_3(x'^{-}), \quad x \in \mathcal{B}_n, \tag{3.49}$$

where $x'^{+} \in \mathcal{S}_n$ is related to $x \in \mathcal{B}_n$ by

$$x_i'^{+} = \begin{cases} x_i, & i = 1, \ldots, n \\ (1 - \sum_{j=1}^{n} x_j^2)^{\frac{1}{2}}, & i = n+1 \end{cases},$$

and where $x'^{-} \in \mathcal{S}_n$ is related to $x \in \mathcal{B}_n$ by

$$x_i'^{-} = \begin{cases} x_i, & i = 1, \ldots, n \\ -(1 - \sum_{j=1}^{n} x_j^2)^{\frac{1}{2}}, & i = n+1 \end{cases}.$$

The proof of Theorem 3 of [15] shows that υ_2 is an ℓ^{th} degree polynomial on $\mathcal{B}_n$ satisfying (3.46). Therefore using (3.49), (3.47), and (3.40), we can construct a polynomial N satisfying (3.41) from a polynomial υ_3 that satisfies (3.48).

A polynomial υ_3 that satisfies (3.48) may be found as follows. Restrict ℓ to be even, and let $p = \frac{\ell}{2}$. For each positive integer i, let $q_i : [-1, 1] \to \mathbb{R}$ be the i^{th}

[15] In [15], in the notation of [15], A_7 may be taken to be $2A_6$, because the number of dimensions ≥ 1, and because a modulus of continuity must be sub-additive and monotonically increasing.

orthogonal polynomial[16] determined by the weight function

$$\left(1-t^2\right)^{\frac{n-2}{2}}, \quad t \in [-1,1],$$

and let λ_i be the largest root of q_i. Let $Q : \mathbb{R} \to \mathbb{R}$ be the polynomial given by

$$Q(t) = \left(\frac{q_{p+1}(t)}{t-\lambda_{p+1}}\right)^2, \quad t \in [-1,1].$$

Also define $\bar{Q}$ by[17]

$$\bar{Q} = \frac{(4\pi)^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}{(n-2)!}\int_{-1}^{1} Q(t)\left(1-t^2\right)^{\frac{n-2}{2}}dt, \tag{3.50}$$

where Γ is the complete Gamma function. Then the proof of Theorem 2 of [15] shows that $\hat{\upsilon}_3 : \mathcal{S}_n \to \mathbb{R}$ given by

$$\hat{\upsilon}_3(x') = \frac{\int_{\mathcal{S}_n} Q\big(\langle x', z'\rangle\big) N_3(z') d\varsigma(z')}{\bar{Q}}, \tag{3.51}$$

where ς denotes hyper-surface measure on $\mathcal{S}_n$ and $\langle\cdot,\cdot\rangle$ denotes the inner product, is an ℓ^{th} degree polynomial satisfying (3.48), with $A = 2(1+\pi\sqrt{10})$.[18]

We may not be able to determine the polynomial $\hat{\upsilon}_3$ analytically from (3.51), because we would need an expression for (3.14) which allows the integral in (3.51)

---

[16] See [31] for material on orthogonal polynomials. The polynomial q_i is the i^{th} degree Jacobi polynomial with "α" = "β" = $\frac{n-2}{2}$, in the notation of [31].

[17] The value of the constant preceding the integral on the right side of (3.50) is not given in [15]. Instead, Lemma 6 of [15] directs the reader to [32]. The value of the constant may be found using the Funk-Hecke Theorem given in Section 11.4 of [32].

[18] Lemma 5 of [15] contains an unspecified constant "a", which is shown in its proof to be $\frac{\pi}{\sqrt{2}}$ for the case in which n is odd (i.e., the case in which $n+1$, equal to "k" in the notation of Theorem 2 and Lemma 5 of [15], is even). In the case of even n, the reader of the proof of Lemma 5 of [15] is directed to use an argument similar the the case of odd n, but with material on Legendre polynomials found in [31]. Using Section 2.4 and Theorem 6.3.2 of [31], it can be seen that "a" can also be $\frac{\pi}{\sqrt{2}}$ when n is even. Because of this, the value of A we have given is valid when n is any positive integer. Note that at the end of the proof of Lemma 5 of [15], a factor of π is erroneously left out of the expression for the zeros of the Chebyshev polynomial of the first kind (see Section 2.4 of [31] or Section 10.11 of [32], for example). The lemma itself is correct as stated.

to have a closed form. Such a requirement might be difficult to meet in practice. Instead, in the next paragraph, we approximate the integral in (3.51) using a quadrature sum. Each term in the quadrature sum depends on the value of $N_3(x')$ for a single $x' \in \mathcal{S}_n$. Therefore by (3.42) and (3.44), to calculate any term in the quadrature sum, we need to know the value of $N_1(x)$ for a single $x \in \mathbb{R}^n$. Each of these values of $N_1(x)$ may be found by evaluating $(Gs)(\Delta)$ for a single s, using (3.14). Note that to accommodate this approximation, we need to use a larger value of A, as specified in the following paragraph.

We find such a quadrature sum as follows. For any positive integer ℓ and any real $\zeta > 0$, we may choose an integer J, coefficients $a_1, \ldots a_J$, and points $z'(1) \ldots z'(J)$ on the sphere $\mathcal{S}_n$ such that

$$\left| \int_{\mathcal{S}_n} Q\big(\langle x', z'\rangle\big) N_3(z') d\varsigma(z') - \sum_{j=1}^{J} a_j Q\big(\langle x', z'(j)\rangle\big) N_3(z'(j)) \right| < \frac{\zeta}{\bar{Q}}\, \bar{\omega}_{N_3}\binom{n}{\ell} \qquad (3.52)$$

for every $x' \in \mathcal{S}_n$. This can be accomplished by, for example, expressing the hypersurface integral as an n-fold multiple integral, and approximating each of these integrals by a Riemann sum. Note that the quadrature sum above will be a polynomial in x' of the same degree, ℓ, as (3.51), regardless of the number J of terms in the quadrature sum. Let

$$\upsilon_3(x') = \frac{\sum_{j=1}^{J} a_j Q\big(\langle x', z'(j)\rangle\big) N_3(z'(j))}{\bar{Q}}, \qquad x' \in \mathcal{S}_n. \qquad (3.53)$$

Then using (3.51), (3.52), and (3.53),

$$\left| \hat{\upsilon}_3(x') - \upsilon_3(x') \right| < \zeta \bar{\omega}_{N_3}\binom{n}{\ell}$$

for every $x' \in \mathcal{S}_n$. Using the triangle inequality, υ_3 is a polynomial of degree ℓ

satisfying (3.48) with

$$A = \zeta + 2(1 + \pi\sqrt{10}).$$

As we indicated earlier, we can construct N satisfying (3.19), with A as above, using υ_3.

3.6 Example G: Nonlinear Feedback System

Recall that in order to construct an L-N structure approximation of a dynamic nonlinear system G using the method discussed in Section 3.3.2, we need a working modulus of continuity for G and a working memory modulus for G. Also, we need to be able to calculate the value of the output of G at a particular time due to any input in $L_{\infty e}(\mathbb{R})$. In this section, we show that we can find all of this information for a feedback system satisfying the familiar circle criterion. This system has previously been used [11] as an example to show that a familiar, but complex, system can be approximated using an L-N structure. Therefore this section establishes that the information we need about G may be obtained for an important class of dynamic nonlinear systems.

The system is as follows (see [11] and Figure 3.1). The input r is an element of $L_{\infty e}(\mathbb{R}_+)$, and it is assumed that there is a solution such that e, w, and y also belong to $L_{\infty e}(\mathbb{R}_+)$. (This is a standard assumption, and it is typically satisfied.)

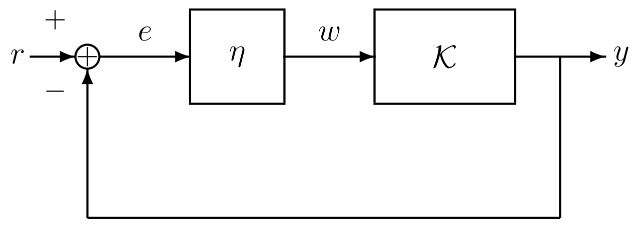

Figure 3.1: Nonlinear feedback system

Let r, e, w, and y be related by

$$\begin{aligned} e(t) &= r(t) - y(t) \\ w(t) &= \eta(e(t)) \qquad\qquad , \qquad t \geq 0 \\ y(t) &= (\mathcal{K}w)(t) + k_0(t) \end{aligned} \tag{3.54}$$

where $k_0 \in L_\infty(\mathbb{R}_+)$ accounts for initial conditions, and where $\mathcal{K}$ and η are as follows. The operator $\mathcal{K}$ is defined by

$$(\mathcal{K}x)(t) = \int_0^t k(t-\tau)x(\tau)d\tau, \qquad t \geq 0$$

for $x \in L_{\infty e}(\mathbb{R}_+)$. We assume $k : \mathbb{R}_+ \to \mathbb{R}$ is the inverse Laplace transform of a rational function

$$K(\sigma) = \frac{\nu(\sigma)}{d(\sigma)}, \qquad \sigma \in \mathbb{C},$$

where the degree of ν is less than the degree of d, ν and d are relatively prime, and the real part of every root of d is no greater than $-\sigma_d$, for some real $\sigma_d > 0$. We further assume that $\eta(0) = 0$, and that for some $\alpha_0, \beta_0 \in \mathbb{R}$ with $\beta_0 > 0$ and $\alpha_0 \leq \beta_0$, the map $\eta : \mathbb{R} \to \mathbb{R}$ satisfies the "sector" condition

$$\alpha_0 \leq \frac{\eta(a_1) - \eta(a_2)}{a_1 - a_2} \leq \beta_0 \tag{3.55}$$

for all real $a_1 \neq a_2$. Assume further that

$$1 + \tfrac{1}{2}(\alpha_0 + \beta_0)K(\sigma) \neq 0 \quad \text{for } \mathrm{Re}(\sigma) \geq 0, \qquad \text{and} \tag{3.56}$$

$$\tfrac{1}{2}(\beta_0 - \alpha_0) \sup_{\omega \in \mathbb{R}} \left| \frac{K(i\omega)}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega)} \right| < 1, \tag{3.57}$$

where $i = \sqrt{-1}$. It is known [33] that the two conditions above are met if one of the following three conditions is satisfied:

1) $0 < \alpha_0 < \beta_0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ lies outside C_1 and does not encircle C_1, where C_1 is the circle (in the complex plane) of radius $\frac{1}{2}(\alpha_0^{-1} - \beta_0^{-1})$ whose center has real part $-\frac{1}{2}(\alpha_0^{-1} + \beta_0^{-1})$ and complex part 0.

2) $0 = \alpha_0 < \beta_0$, and the real part of $K(i\omega)$ is greater than $-\beta_0^{-1}$ for all real ω.

3) $\alpha_0 < 0 < \beta_0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ is contained within C_2, where C_2 is the circle (in the complex plane) of radius $\frac{1}{2}(\beta_0^{-1} - \alpha_0^{-1})$ whose center has real part $-\frac{1}{2}(\alpha_0^{-1} + \beta_0^{-1})$ and complex part 0.

Under these conditions, it is shown in Section III of [11] that when $k_0(t) = 0$ for all $t \in \mathbb{R}_+$, there is a unique map $G : L_{\infty e}(\mathbb{R}_+) \to L_{\infty e}(\mathbb{R}_+)$ taking r to y in accordance with (3.54). It is also shown that G has approximately finite memory and uniform continuity property $\mathcal{P}_c$, and that G is causal and time-invariant.[19]

3.6.1 Finding a Working Modulus of Continuity $\bar{\mu}_G$

In this section we determine a working modulus of continuity $\bar{\mu}_G$ for G. We can see that G has uniform continuity property $\mathcal{P}_c$ from the proof of Theorem 4 of [11]. If we follow the reference in [11] to [33], and in [33] to [34], we can also see how to find $\bar{\mu}_G$. From this material, we see that for every $s_1, s_2 \in L_2(\mathbb{R}_+)$ and every $t \in \mathbb{R}_+$,

$$|(Gs_1)(t) - (Gs_2)(t)| \leq \rho \|s_1 - s_2\|_2,$$

[19]It should be noted that G describes the system even in the case of nonzero initial conditions. Specifically, if the condition $k_0(t) = 0$ is not met, the relationship between the input r and the output y still depends on G according to the following equation:

$$y(t) = (G[r - k_0])(t) + k_0(t).$$

So, results shown to hold for G are not limited in scope to the case of zero initial conditions.

where

$$\rho = \rho' \max(\beta_0, -\alpha_0)\left(\int_0^\infty |k(t)|^2 dt\right)^{\frac{1}{2}}$$
$$\rho' = \frac{\kappa_1}{1 - \frac{1}{2}\kappa_2(\beta_0 - \alpha_0)}$$
$$\kappa_1 = \sup_{\omega\in\mathbb{R}}\left\{\left|\frac{1}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega)}\right|\right\}$$
$$\kappa_2 = \sup_{\omega\in\mathbb{R}}\left\{\left|\frac{K(i\omega)}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega)}\right|\right\}.$$

(It is shown in Section IV of [34] that $\frac{1}{2}|\alpha_0 - \beta_0|\kappa_2 < 1$, so that ρ is always finite and positive.) Therefore,

$$\bar{\mu}_G(\delta) = \rho\delta, \quad \delta > 0 \tag{3.58}$$

is a working modulus of continuity.

3.6.2 Finding a Working Memory Modulus $\bar{m}_G$

In this section we determine a working memory modulus $\bar{m}_G$ for G. Theorem 4 of [11] says that G has approximately finite memory. We can find $\bar{m}_G$ by following the proof, which leads us to [26], which in turn leads us to [34]. The result is as follows.

Assume $k_0(t) = 0$ for all $t \in \mathbb{R}_+$. Choose a real number $\sigma_0 \in (0, \sigma_d)$ such that

$$\max(\beta_0, -\alpha_0) \sup_{\omega\in\mathbb{R}} |K(i\omega - \sigma_0)| < 1.$$

(Lemma 3 of [26] shows that there are values of σ_0 which satisfy this restriction.) Then for every $\Delta > 0$ and every $\tau \in \mathbb{R}_+$,

$$|(G_1 s)(\tau) - (G_1 W_{\tau,\Delta} s)(\tau)| \le \frac{\lambda b}{(2\sigma_0)^{\frac{1}{2}}} e^{-\sigma_0\Delta}, \tag{3.59}$$

where

$$\lambda = \lambda' \left(\int_0^\infty |e^{\sigma_0 t} k(t)|^2 dt \right)^{\frac{1}{2}} \max(\beta_0, -\alpha_0)$$
$$\lambda' = \frac{\kappa_3}{1 - \frac{1}{2}\kappa_4(\beta_0 - \alpha_0)}$$
$$\kappa_3 = \sup_{\omega\in\mathbb{R}} \left\{ \left| \frac{1}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega - \sigma_0)} \right| \right\}$$
$$\kappa_4 = \sup_{\omega\in\mathbb{R}} \left\{ \left| \frac{K(i\omega - \sigma_0)}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega - \sigma_0)} \right| \right\}.$$

Therefore,

$$\bar{m}_G(\Delta) = \frac{\lambda b}{(2\sigma_0)^{\frac{1}{2}}} e^{-\sigma_0 \Delta}$$

is a working modulus of continuity.

3.6.3 Calculating $(Gs)(\Delta)$

As we have discussed, in order to construct an L-N structure approximation for G, we need to be able to calculate the value of $(Gs)(\Delta)$ for certain inputs $s \in L_{\infty e}(\mathbb{R}_+)$. Specifically, to find the nonlinear map $N \in \Upsilon_n$, we need to evaluate $N_1(z)$ for a finite number of points $z \in \mathbb{R}^n$, as indicated in Section 3.5.1 (if Υ_n is the set of Gaussian radial basis functions) or in Section 3.5.2 (if Υ_n is the set of polynomials).[20] Each $N_1(z)$ may be found by evaluating $(Gs)(\Delta)$ for a certain value of $s \in L_{\infty e}(\mathbb{R}_+)$, as shown in (3.14). Because it is not immediately obvious how one might calculate $(Gs)(\Delta)$ in our feedback system example, we give the following iterative technique based on the principle of Contraction Mapping.[21]

When $k_0(t) = 0$ for all $t \in \mathbb{R}_+$, the mapping which takes r of equations (3.54) into e may be expressed as

$$r = e + \mathcal{K}\hat{\eta}e, \tag{3.60}$$

[20] See equation (3.36) or (3.53), respectively.

[21] One discussion of the principle of Contraction Mapping may be found in §7 of [29].

where $\hat{\eta}$ is the memoryless map from $L_{\infty e}(\mathbb{R}_+)$ to $L_{\infty e}(\mathbb{R}_+)$ given by $(\hat{\eta}s)(t) = \eta(s(t))$ for each $s \in L_{\infty e}(\mathbb{R}_+)$ and each $t \geq 0$. Adding and subtracting $\mathcal{K}(\frac{1}{2}(\alpha_0 + \beta_0)e)$ on the right side of (3.60), and using the linearity of $\mathcal{K}$, we have

$$\begin{aligned} r &= e + \mathcal{K}(\tfrac{1}{2}(\alpha_0 + \beta_0)e) + \mathcal{K}(\hat{\eta} - \tfrac{1}{2}(\alpha_0 + \beta_0)I)e \\ &= (I + \tfrac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})e + \mathcal{K}(\hat{\eta} - \tfrac{1}{2}(\alpha_0 + \beta_0)I)e, \end{aligned} \tag{3.61}$$

where I is the identity map on $L_{\infty e}(\mathbb{R}_+)$. Now from (3.56), we see that $(I + \frac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})$ has a causal inverse given by

$$\big((I + \tfrac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})^{-1}s\big)(t) = \int_0^t h(t - \tau)s(\tau)d\tau, \qquad s \in L_{\infty e}(\mathbb{R}_+),\ t \in \mathbb{R}_+,$$

where $h \in L_1(\mathbb{R}_+)$ is the inverse Laplace transform of the rational function

$$\frac{1}{1 + \frac{1}{2}(\alpha_0 + \beta_0)K(\sigma)} \tag{3.62}$$

on a region of convergence including all $\sigma \in \mathbb{C}$ with nonnegative real parts. Therefore (3.61) is equivalent to

$$\begin{aligned} e = {} & (I + \tfrac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})^{-1}r \\ & - (I + \tfrac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})^{-1}\mathcal{K}(\hat{\eta} - \tfrac{1}{2}(\alpha_0 + \beta_0)I)e. \end{aligned} \tag{3.63}$$

Consider the right side of (3.63) above. Using the sector condition of (3.55), and recalling that $\beta_0 > \alpha_0$, we can see that for any $a_1, a_2 \in \mathbb{R}$

$$\big|\eta(a_1) - \eta(a_2) - \tfrac{1}{2}(\alpha_0 + \beta_0)\big| \leq \tfrac{1}{2}(\beta_0 - \alpha_0)(a_1 - a_2). \tag{3.64}$$

Therefore with $a_2 = 0$, we see that if e is in $L_2(\mathbb{R}_+)$, then so is $(\hat{\eta} - \frac{1}{2}(\alpha_0 + \beta_0)I)e$. Further, since the convolution kernels k and h are inverse Laplace transforms of rational functions with poles confined to the left half plane, we have $k, h \in L_1(\mathbb{R}_+)$.

So by Young's Convolution Theorem (see Theorem 9.2 of [28], for example), if e and r are both in $L_2(\mathbb{R}_+)$, then so are both terms on the right side of equation (3.63).

Now fix $r \in L_2(\mathbb{R}_+)$, and define $A_{\Delta,r} : L_2(\mathbb{R}_+) \to L_2(\mathbb{R}_+)$ by

$$\begin{aligned} A_{\Delta,r}x = W_{\Delta,\Delta}\Big(& (I + \tfrac{1}{2}(\alpha_0+\beta_0)\mathcal{K})^{-1} r \\ & - (I + \tfrac{1}{2}(\alpha_0+\beta_0)\mathcal{K})^{-1}\mathcal{K}(\hat{\eta} - \tfrac{1}{2}(\alpha_0+\beta_0)I)W_{\Delta,\Delta}x\Big) \end{aligned} \tag{3.65}$$

for each $x \in L_2(\mathbb{R}_+)$. If $x, x' \in L_2(\mathbb{R}_+)$, then by Lemma 5 of [34],

$$\begin{aligned} \|A_{\Delta,r}x - & A_{\Delta,r}x'\|_2 \\ & \le \big\|(I + \tfrac{1}{2}(\alpha_0+\beta_0)\mathcal{K})^{-1}\mathcal{K}\big(\hat{\eta} - \tfrac{1}{2}(\alpha_0+\beta_0)I\big)W_{\Delta,\Delta}(x - x')\big\|_2 \\ & \le \sup_{\omega\in\mathbb{R}}\left|\frac{K(i\omega)}{1 + \frac{1}{2}(\alpha_0+\beta_0)K(i\omega)}\right| \big\|\big(\hat{\eta} - \tfrac{1}{2}(\alpha_0+\beta_0)I\big)W_{\Delta,\Delta}(x - x')\big\|_2 \end{aligned}$$

Further, it follows easily from (3.64) that

$$\big\|\big(\hat{\eta} - \tfrac{1}{2}(\alpha_0+\beta_0)I\big)W_{\Delta,\Delta}(x - x')\big\|_2 \le \tfrac{1}{2}(\beta_0 - \alpha_0)\|x - x'\|_2.$$

Therefore

$$\big\|A_{\Delta,r}x - A_{\Delta,r}x'\|_2 \le \phi\|x - x'\|_2,$$

where

$$\phi = \tfrac{1}{2}(\beta_0 - \alpha_0)\sup_{\omega\in\mathbb{R}}\left|\frac{K(i\omega)}{1 + \frac{1}{2}(\alpha_0+\beta_0)K(i\omega)}\right|.$$

Recall from (3.57) that $\phi < 1$. So $A_{\Delta,r}$ satisfies the conditions of the Contraction Mapping Theorem (see §7 of [29], for example).

Since $A_{\Delta,r}$ is a contraction map, we can proceed as follows. Define $A^1_{\Delta,r} = A_{\Delta,r}$, and for each integer $j \ge 2$, define $A^j_{\Delta,r} : L_2(\mathbb{R}_+) \to L_2(\mathbb{R}_+)$ by

$$A^j_{\Delta,r}x = A_{\Delta,r}A^{j-1}_{\Delta,r}x, \quad x \in L_2(\mathbb{R}_+).$$

Let x_0 be any element of $L_2(\mathbb{R}_+)$. Using the principle of Contraction Mapping (as in §7 of [29]), there is a unique x_r such that $x_r = A_{\Delta,r}x_r$, and furthermore, for every positive integer j,

$$\|A^j_{\Delta,r}x_0 - x_r\|_2 \leq \frac{\phi^j}{1-\phi}\|x_0 - A_{\Delta,r}x_0\|_2. \tag{3.66}$$

Now choose any $s \in L_{\infty e}(\mathbb{R}_+)$. Set $r = W_{\Delta,\Delta}s$, and note that $r \in L_2(\mathbb{R}_+)$. Let e be the unique element of $L_{\infty e}$ related to r by (3.54), and observe that $W_{\Delta,\Delta}e \in L_2(\mathbb{R}_+)$. Choose any $x_0 \in L_2(\mathbb{R}_+)$. Assume $x_r = A_{\Delta,r}x_r$ is as above. Because $\mathcal{K}$ and $(I + \frac{1}{2}(\alpha_0 + \beta_0)\mathcal{K})^{-1}$ are causal, we see from (3.63) and (3.65) that

$$W_{\Delta,\Delta}e = A_{\Delta,r}W_{\Delta,\Delta}e.$$

But since x_r is the only element of $L_2(\mathbb{R}_+)$ which satisfies $A_{\Delta,r}x_r = x_r$,

$$W_{\Delta,\Delta}e = x_r.$$

Using the causality of $\mathcal{K}$, we conclude that

$$(Gs)(\Delta) = (\mathcal{K}\hat{\eta}e)(\Delta) = (\mathcal{K}\hat{\eta}W_{\Delta,\Delta}e)(\Delta) = (\mathcal{K}\hat{\eta}x_r)(\Delta). \tag{3.67}$$

Now let $\lambda > 0$, and choose any $x_0 \in L_2(\mathbb{R}_+)$. Pick a positive integer j such that

$$\|k\|_2 \max\{|\beta_0|, |\alpha_0|\}\|x_0 - A_{\Delta,r}x_0\|_2 \,\frac{\phi^j}{1-\phi} < \lambda. \tag{3.68}$$

Using Holder's inequality,

$$\begin{aligned}\big|(\mathcal{K}\hat{\eta}x_r)(\Delta) - (\mathcal{K}\hat{\eta}A^j_{\Delta,r}x_0)(\Delta)\big| &= \left|\int_0^\tau k(t-\tau)\big[\eta(x_r(\tau)) - \eta((A^j_{\Delta,r}x_0)(\tau))\big]\,d\tau\right| \\ &\leq \|k\|_2\left(\int_0^\tau \big|\eta(x_r(\tau)) - \eta((A^j_{\Delta,r}x_0)(\tau))\big|^2 d\tau\right)^{\frac{1}{2}}.\end{aligned}$$

The sector condition (3.55), together with (3.66), gives

$$\begin{aligned}\left|(\mathcal{K}\hat{\eta}x_r)(\Delta) - (\mathcal{K}\hat{\eta}A^j_{\Delta,r}x_0)(\Delta)\right| &\le \|k\|_2 \max\{|\beta_0|, |\alpha_0|\}\, \|x_r - A^j_{\Delta,r}x_0\|_2 \\ &\le \|k\|_2 \max\{|\beta_0|, |\alpha_0|\}\, \|x_0 - A_{\Delta,r}x_0\|_2 \,\frac{\phi^j}{1-\phi},\end{aligned}$$

and therefore by (3.67) and (3.68),

$$\left|(Gs)(\Delta) - (\mathcal{K}\hat{\eta}A^j_{\Delta,r}x_0)(\Delta)\right| \le \lambda.$$

Because the value of $(\mathcal{K}\hat{\eta}A^j_{\Delta,r}x_0)(\Delta)$ can be calculated in a finite number of steps, we can calculate the value of $(Gs)(\Delta)$ as accurately as we like.

In Section 3.5, we did not explicitly consider the effect on N of inexact evaluation of $(Gs)(\Delta)$. However, the additional error introduced is bounded in a satisfactory fashion. In the case of Gaussian radial basis function networks, replacing $(Gs)(\Delta)$ in (3.14) with the approximation $(\mathcal{K}\hat{\eta}A^j_{\Delta,r}x_0)(\Delta)$ as above changes the value of each ξ'_θ in (3.36) by no more than $(\frac{2B}{\ell})^n\lambda$. Because (in the notation of Section 3.5.1) there are ℓ^n terms in the Gaussian radial basis network sum, and because $|\gamma(z)| \le (\frac{1}{\sqrt{\pi}})^n$ for every $z \in \mathbb{R}^n$, this substitution introduces no more than $(\frac{2B}{\sqrt{\pi}})^n\lambda$ of additional error into the approximation of N_1. Therefore the additional error can be made arbitrarily small as above. In the case of polynomial networks, a similar argument bounds the additional error introduced by this substitution into the quadrature sum of (3.52), and into the polynomial approximation of N_1.

3.7 Complexity of Approximations

The results presented in this chapter are important because they are the first constructive results for uniform approximation of dynamic nonlinear systems using L-N structures. As we indicated earlier, we do not claim that the complexity of an L-

N structure constructed as in this chapter is satisfactory. In fact, certain results presented in [15] raise concerns about the required complexity of N. In the case in which Υ_n is composed of polynomials, because the dimension of the set of degree ℓ polynomials in $\mathbb{R}^n$ is $\binom{\ell+n}{n}$, and because ℓ increases with decreasing ε_n according to (3.41), the number of required terms grows very quickly as the error tolerance ε_n is tightened. However, it follows from Theorem 1 of [15] that no choice of Υ_n having $\binom{\ell+n}{n}$ linear dimensions can give a better result for the number of required terms than (3.41), except that the value of the multiplicative constant A may be different. Therefore, approximation structures formed by superpositions of fixed simple elements will give approximation bounds for which the number of required terms grows at least as quickly as for polynomials in (3.41), as the error tolerance ε_n is tightened.[22]

We note that better uniform error bounds have been discovered for the case in which N_1 satisfies additional regularity conditions involving, for example, integral representations [35], a bound on a certain kind of variation [36], or a bound on a spectral norm [36]. Use of these improved error bounds is therefore limited to special cases of G. However, it may be difficult to find reasonable restrictions on G which result in N_1 of (3.14) satisfying these regularity conditions.

We note that well-known results exist regarding better error bounds for mean square approximation using sigmoidal neural networks (see Chapter 25 of [18], [37], and [38]). However, a bound on the mean square error of N does not permit us to guarantee uniform approximation for L-N structures, so these results do not help to address the problem examined in this chapter.

[22]The results in [15] do not place restrictions on approximation techniques in which nonlinear parameters are adjusted. Therefore the restrictions do not apply to approximation techniques using radial basis function networks and sigmoidal neural networks, if, for example, the internal dilation, orientation, or translation parameters are allowed to vary continuously during the approximation process. However, this author is not aware of any improved uniform error bounds of this type, for the general case.

Chapter 4

Myopic L_p–Compactness

4.1 Introduction

As mentioned in Chapter 1, Nonlinear filters play an important role in image processing. In a common scenario, one would like to enhance an image by removing non-Gaussian noise such as "shot" noise, while leaving feature edges intact. Noise and feature edges are difficult to separate using linear filters, because both are high-frequency components of the image [5]. Other applications of nonlinear filters in image processing include edge detection, image sharpening, and image compression ([5], [39]).

A host of nonlinear filter structures have arisen in the literature of image processing, as attempts have been made to satisfactorily address these issues. Such filter structures include median filters, stack filters, rational filters, partial differential equations, and order-statistic filters ([5], [39]). New filter structures and variations on previous structures are still being studied ([40], [41]), because there is still a need for improved filter structures.

The simplicity of the L-N structure and its powerful approximation properties show great potential for applications in image processing. One particular result

[12] shows that arbitrarily accurate L-N structure approximations exist for a broad class of nonlinear mappings from input functions to output functions. The domain of the input and output functions in [12] is in $\mathbb{R}^m$, where m can be any positive integer. The range of the input functions is in $\mathbb{R}^\ell$, where ℓ can be any positive integer, and the range of the output functions is in $\mathbb{R}$. This is ideal for image processing, because many images may be represented by functions whose domain is in $\mathbb{R}^2$, or possibly $\mathbb{R}^3$ (for three-dimensional images [24] or moving images [25]), and whose range is in $\mathbb{R}$ (for gray-scale images) or $\mathbb{R}^3$ (for color images).[1] Further, the input functions in [12] are permitted to have discontinuities. This is important because functions representing images typically have discontinuities formed by feature edges. The approximation property central to [12] is interesting in the context of image processing because a similar approximation property for rational filters has already proved to be important [42].

For the L-N structure approximation result of [12] to hold for a given nonlinear map, the map must be myopic, in a certain sense. Roughly, this means that the value of the output signal of the map, at any point x, is changed little by variations in values of the input signal at points far from x, or by small variations in values of the input signal at points near x. This condition is both reasonable and easy to understand. However, the input set over which the approximation result is valid must meet a certain compactness condition which is stated in terms of the topology of the input set, rather than in terms of more tangible criteria.[2] It is clear from the Arzela-Ascoli theorem that a set of continuous functions with a uniform magnitude bound, and with a uniform bound on the slope of all the functions at all points, would satisfy the requirements of the theorem. Although this is a tangible criterion, images typically have discontinuities caused by feature edges. These feature edges

[1] An $\mathbb{R}^3$-valued output function may easily be viewed as three distinct $\mathbb{R}$-valued functions which are the outputs of three distinct nonlinear mappings.

[2] The input set must also be uniformly bounded in magnitude. This boundedness condition does not present any problems in terms of tangibility or practicality.

are often key components of the image, so an continuity condition would severely limit practical applications in image processing. Therefore it is not clear from [12] whether one can define an input set in terms of criteria which are tangible and practical in the context of image processing.

In the discrete-time case [43], this issue does not arise, because a uniform magnitude bound alone is sufficient. In the continuous-time case in which no discontinuities are allowed in the input set, an equicontinuity condition is all that is needed [44]. In the case $m = 1$ in [11], a compactness condition similar to that in [12] is placed on the input set. There, an interesting set is given which includes elements with discontinuities, and which satisfies the compactness condition. However, no such set has previously been given for the case of [12] when $m > 1$.

In this chapter we give sets which satisfy the compactness criterion of [12], and which are tangible and practical in the context of image processing. The sets include elements containing discontinuities. This allows us to establish that the results in [12] can be applied to sets of input functions representing interesting images. We will refer to sets which satisfy the compactness criterion in [12] as being "Myopically L_p–Compact," where p may be any real number not less than 1.

In this chapter, we think of images as "analog" entities, i.e., maps whose domain is $\mathbb{R}^2$ or $\mathbb{R}^3$. This is a departure from much of the image processing literature, in which an image is thought of as a digital map of values on a discrete, square grid. We do not mean to imply that we expect image filtering to be performed on analog representations of images, or implemented by analog means. Rather we take the perspective that because the real world is analog, a digital image is a representation of an analog image. Therefore in this chapter we address analog operations on analog images, and consider the issue of digitization to be an important issue which falls outside the scope of this chapter.[3] This perspective is becoming increasingly

[3]For a recent result in the area of image sampling, and a bibliography of earlier work, see [45].

important as digital camera pixel densities and computer data storage capacities increase without accompanying increases in camera lens resolution, so that analog image limitations often outweigh the limitations of digitization.

Another reason for representing images in analog form is that not all digital images are represented on a square grid. For example, digital radar and sonar images form a grid in which the discrete samples are evenly spaced in range and angle, relative to the transducer. Such digital images may be represented as analog maps by subdividing $\mathbb{R}^2$ along circles centered at the transducer, and along radial lines emanating from the transducer. In each region, the value of the analog map may be set to the value of a digital sample in the center of the region. Therefore image processing techniques developed in an analog context may readily be applied to digital radar and sonar images. By contrast, image processing techniques developed for discrete sampling on square grids cannot be directly applied to radar or sonar images without introducing spatial distortion.

A final motivation for considering analog images is that a large number of computer-generated images are defined not by a digital map of values on a discrete grid (i.e., a rasterized image consisting of pixels), but in terms of a collection of parameterized geometric shapes. For example, scalable fonts, such as Adobe Postscript Type 1 fonts, represent each character by a geometric image. Many modern drawing programs, such as Adobe Illustrator, Flash, and most Postscript drawing programs, use "vector graphics" to describe the images using geometric shapes.[4] Most three-dimensional graphics systems use vertices of polygons to describe a computer-generated scene, so that a two-dimensional image of the scene viewed from a particular point is described by the three-dimensional geometry, rather than by pixels. Therefore representation of images as functions defined on $\mathbb{R}^2$, instead of on a discrete square grid, is important in the context of computer-generated images.

[4]Many of these programs have provisions for importing rasterized content, but content created within the program is typically described using vector graphics.

This chapter is organized as follows. In Section 4.2, we state the compactness condition which is the topic of this chapter. In each subsequent section, we state, discuss, and prove a theorem that defines a class of tangible sets satisfying the compactness condition. Theorem 4.1 of Section 4.3 is concerned with counting the number of times certain line segments cross feature edges. A Lipschitz condition is also involved. Theorem 4.2 of Section 4.4 is concerned with total variation of the function representing the image, as measured along certain line segments. We show that Theorem 4.2 may be viewed as a generalization of Theorem 4.1. This observation allows feature edges that do not form true discontinuities to be counted as discontinuities in a feature edge counting process similar to that of Theorem 4.1. Theorem 4.3 of Section 4.5 is concerned with the number and total arc length of feature edges within squares of a certain size. Again, a Lipschitz condition is also involved.

4.2 Myopically L_p–Compact Sets

As we stated earlier, the approximation results in [12] are shown to be valid on input sets satisfying a certain criterion involving compactness. In this section we give this compactness criterion. We use the term "myopically L_p–compact" throughout this chapter to describe sets which satisfy the criterion.[5]

First we need the following definitions. For each positive integer k, let $|\cdot|_1$, $|\cdot|_2$, and $|\cdot|_\infty$ denote the norms on $\mathbb{R}^k$ given for each $x \in \mathbb{R}^k$ by

$$|x|_1 = \sum_{i=1}^{k} |x_i|, \qquad |x|_2 = \left(\sum_{i=1}^{k} x_i^2\right)^{\frac{1}{2}}, \qquad |x|_\infty = \max_{i \in \{1,\dots,k\}} |x_i|.$$

For each positive integer k, each $r > 0$ and each $x \in \mathbb{R}^k$, let $B_r(x)$ denote the open

[5] The term "myopically L_p–compact" is not used in [12]. We introduce it here because we refer to the criterion repeatedly.

ball in $\mathbb{R}^k$ of radius r centered at x, i.e., $B_r(x) = \{y \in \mathbb{R}^k : |x-y|_2 < r\}$, and let $\bar{B}_r(x)$ denote the closed ball in $\mathbb{R}^k$ of radius r centered at x, i.e., $\bar{B}_r(x) = \{y \in \mathbb{R}^k : |x-y|_2 \leq r\}$.

Let ℓ and m be positive integers. For each $p \geq 1$ and each Lebesgue measurable $D \subseteq \mathbb{R}^m$, let $L_p^\ell(D)$ denote the set of all measurable $s : \mathbb{R}^m \to \mathbb{R}^\ell$ such that

$$\int_D |s(x)|_1^p dx < \infty,$$

and for each $s \in L_p^\ell(D)$, let $\|s\|_p$ denote the norm[6]

$$\|s\|_p = \left(\int_D |s(x)|_1^p dx\right)^{\frac{1}{p}}.$$

As is customary, to avoid cumbersome notation, we will not explicitly differentiate between $L_p^\ell(\mathbb{R}^m)$ as set of functions and $L_p^\ell(\mathbb{R}^m)$ as a metric space. It is well-known that these two are at technically at odds, in the sense that two functions which differ only on a set of Lebesgue measure zero are equivalent elements of the metric space. We raise the issue here because the conditions of the theorems in this chapter are stated in terms of functions, while the conclusions state that certain sets of functions have a certain metric space topological property. (This property is given in the following paragraph.) In fact, one function may satisfy the conditions of Theorem 4.1, 4.2, or 4.3, while another function which differs from that function only on a set of measure zero may not satisfy the same conditions. For this reason we clarify that when we say a set of functions is relatively compact in $L_p^\ell(\mathbb{R}^m)$, we mean that the set of metric space elements corresponding to those functions is relatively compact in the metric space $L_p^\ell(\mathbb{R}^m)$.

[6]The norm $\|\cdot\|_p$ on $L_p^\ell(D)$ may instead be defined with $|\cdot|_2$, $|\cdot|_\infty$, or some other norm on $\mathbb{R}^\ell$ in place of $|\cdot|_1$. Using the fact that all norms on $\mathbb{R}^\ell$ are topologically equivalent, it is easy to see that these definitions of $\|\cdot\|_p$ are all topologically equivalent. We are therefore free to use topological results from other sources that define $\|\cdot\|_p$ using other norms on $\mathbb{R}^\ell$, and we will do so without further comment.

The compactness criterion from [12] is as follows. Let $p \geq 1$. A set S of Lebesgue-measurable maps $s : \mathbb{R}^m \to \mathbb{R}^\ell$ is said to be "myopically L_p–compact" if all three of the following conditions are satisfied:

(i) For some $b > 0$, $|s(x)|_2 \leq b$ for every every $s \in S$ and every $x \in \mathbb{R}^m$.

(ii) For every $r > 0$, the set $\{s|_{\bar{B}_r(0)} : s \in S\}$ of restrictions of elements of S to $\bar{B}_r(0)$ is relatively compact[7] as a subset of the metric space $L_p^\ell(\bar{B}_r(0))$.

(iii) For every $s \in S$ and every $x \in \mathbb{R}^m$, we have $s(\cdot - x) \in S$ (i.e. S is "translation-invariant").

If S is a set of functions representing images, then the meanings of conditions (i) and (iii) above are quite straightforward. However, condition (ii) describes a condition on S in terms of topology, and not in terms of features which we may identify in an image. Theorems 4.1, 4.2, and 4.3 in the following sections describe myopically L_p–compact sets in terms of image features.

4.3 Myopically L_p–Compact Sets and Crossings of Feature Edges

Theorem 4.1 of this section gives a myopically L_p–compact set of functions characterized using a bound on the number of function discontinuities (such as image feature edges) crossed by certain line segments, and a Lipschitz condition on regions of continuity along the same line segments. This theorem is closely related to Theorem 3 of [11], which is concerned with the simpler case in which the input signals are defined on $\mathbb{R}_+$.

We need the following definitions. For each Lebesgue measurable $D \subseteq \mathbb{R}^m$,

[7] When we say that a set is relatively compact, we mean that its closure is compact.

let $L^\ell_\infty(D)$ be the set of all measurable $s : \mathbb{R}^m \to \mathbb{R}^\ell$ such that

$$\sup_{x \in D} |s(x)|_\infty < \infty,$$

and for each $s \in L^\ell_\infty(D)$, let $\|s\|_\infty$ denote the norm[8]

$$\|s\|_\infty = \sup_{x \in D} |s(x)|_\infty.$$

For each $t \in \mathbb{R}$, each $i = 1, \ldots, m$, and each $y \in \mathbb{R}^{m-1}$, let $\rho(t, i, y)$ be the element of $\mathbb{R}^m$ whose i^{th} component is t and whose other components are the components of y, i.e., the j^{th} component of $\rho(t, i, y)$ is

$$(\rho(t, i, y))_j = \begin{cases} y_j, & j < i \\ t, & j = i \\ y_{j-1}, & j > i \end{cases}.$$

For each positive integer k, let 0_k be the element of $\mathbb{R}^k$ for which each component is zero.

Theorem 4.1: Let $b, \gamma > 0$, and let $\lambda, \kappa \geq 0$, with the restriction that κ must be an integer. Define $S_1(b, \gamma, \lambda, \kappa)$ as the set of all $s \in L^\ell_\infty(\mathbb{R}^m)$ such that

(i) $\|s\|_\infty \leq b$,

(ii) for every $x \in \mathbb{R}^m$ and every $i \in \{1, \ldots, m\}$, the map from $[0, \gamma]$ to $\mathbb{R}^\ell$ given by

$$s(x + \rho(\tau, i, 0_{m-1})), \qquad \tau \in [0, \gamma], \tag{4.1}$$

[8] As was the case for $L^\ell_p(D)$, the norm $\|\cdot\|_\infty$ on $L^\ell_\infty(D)$ may also defined with $|\cdot|_1$, $|\cdot|_2$, or some other norm on $\mathbb{R}^\ell$ in place of $|\cdot|_\infty$. Using the fact that all norms on $\mathbb{R}^\ell$ are topologically equivalent, it is easy to see that these other definitions of $\|\cdot\|_\infty$ are topologically equivalent. We are therefore free to use topological results from other sources that define $\|\cdot\|_\infty$ using other norms on $\mathbb{R}^\ell$, and we will do so without further comment.

has no more than κ points of discontinuity, and

(iii) on every interval of continuity of each map (4.1) above, each component of the map (4.1) is Lipschitz continuous, with Lipschitz constant λ. (By this we mean that for every $x \in \mathbb{R}^m$, every $i \in \{1, \ldots, m\}$, and every $\tau, \tau' \in [0, \gamma]$ with $\tau > \tau'$,

$$\left| s(x + \rho(\tau, i, 0_{m-1})) - s(x + \rho(\tau', i, 0_{m-1})) \right| \leq \lambda(\tau - \tau'),$$

provided that $s(x + \rho(\cdot, i, 0_{m-1}))$ contains no discontinuities on the closed interval $[\tau', \tau]$.)

then $S_1(b, \gamma, \lambda, \kappa)$ is myopically L_p–compact for every $p \geq 1$.

It is shown in the discussion of Theorem 4.2 of next section that Theorem 4.1 follows immediately from Theorem 4.2. Therefore we do not prove Theorem 4.1 here.

Condition (i) is typically satisfied for a two-dimensional visual image if b represents the brightest value in the image. We may determine whether (ii) is satisfied by moving a ruler horizontally and vertically over an image to see whether more than κ feature edges are crossed within a distance γ of one another. The image satisfies (iii) if the largest rate of change in image intensity along the edge of the ruler (as above) on a region without discontinuities is bounded by λ.

Consider the gray-scale image of an origami crane shown in Figure 4.1. The function representing this image belongs to $S_1(b, \gamma, \lambda, \kappa)$ if we choose b, γ, λ, and κ as follows. Choose b to be no less than the value corresponding to the brightest value in the image, which occurs inside the fold on the front of the hump between the wings of the crane. The shortest line segment intersecting 9 feature edges (including edges in shadows) occurs across the beak and neck of the crane, as pictured in Figure 4.2. Therefore we may choose any $\kappa \geq 8$ and any γ shorter than the line segment in Figure 4.2.[9] The largest horizontal or vertical rate of change in image intensity occurs horizontally in the shadow beneath the near wing of the crane. So we may choose any λ which is no less than the maximum horizontal rate of change in the image intensity on a line segment crossing this shadow.

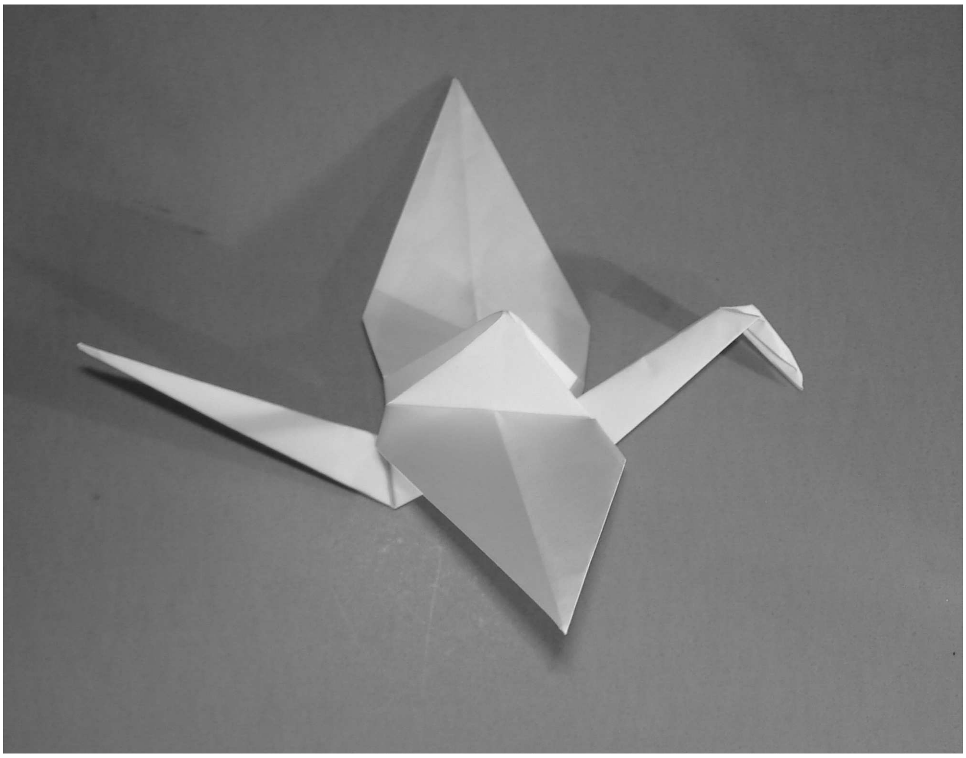

Figure 4.1: An image of an origami crane

[9] There is nothing special about 8 and 9, but to prevent γ from approaching zero, we may not allow κ to be less than the maximum number of feature edges that come together at any single point from any half-plane. For example, 3 discontinuities come together at the near wingtip of the crane. So for $S_1(b, \gamma, \lambda, \kappa)$ to contain this image, κ cannot be less that 3.

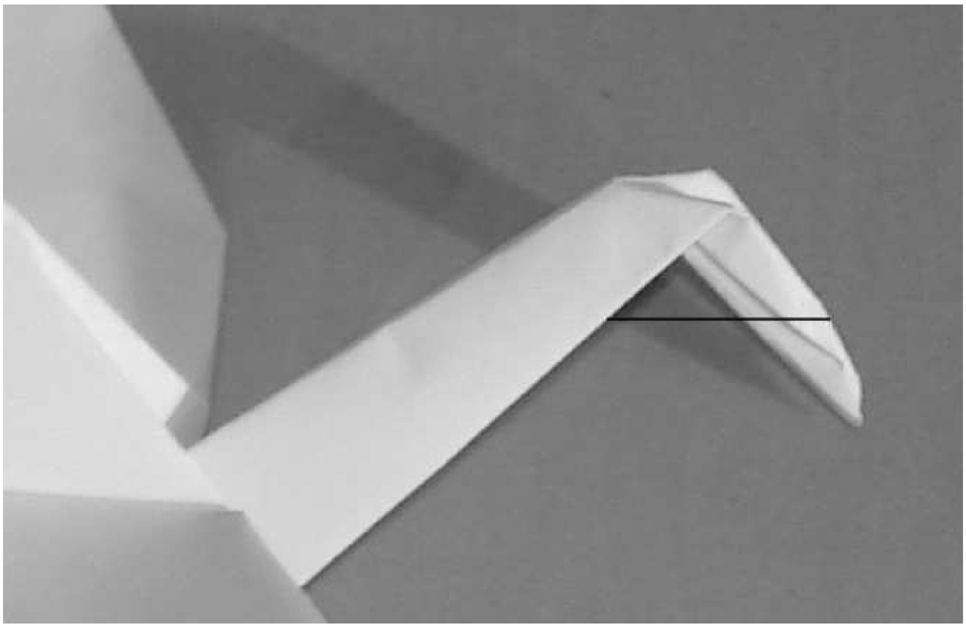

Figure 4.2: Detail of the neck and head of the crane of Figure 4.1

4.4 Myopically L_p–Compact Sets and Total Variation

4.4.1 Theorem and Discussion

Theorem 4.2 of this section is an important generalization of Theorem 4.1. It gives a myopically L_p–compact set of functions characterized by uniform bounds on the magnitude of the functions and on the variation of the functions along certain line segments. It is shown below that Theorem 4.2 allows us to apply conditions similar to those of Theorem 4.1 even if feature edges are not perfectly discontinuous.

We need the following for the statement and proof of Theorem 4.2. For any closed interval $[a, b] \subseteq \mathbb{R}$, a partition of $[a, b]$ is a finite set $\{p_0, \ldots, p_k\}$ of points in $[a, b]$, where $p_0 = a$, $p_k = b$, and $p_{j-1} < p_j$ for $j = 1, \ldots, k$. Let $\mathcal{P}(a, b)$ be the set of all partitions of $[a, b]$. A function $f : [a, b] \to \mathbb{R}$ is said to have bounded variation on $[a, b]$ if

$$\sup_{P \in \mathcal{P}(a,b)} \sum_{j=1}^{k_P} \bigl| f(p_j) - f(p_{j-1}) \bigr| \tag{4.2}$$

is finite (where $\{p_0, \ldots, p_{k_P}\} = P$). In this case, we say that (4.2) is the total variation of f on $[a, b]$. For each $w : [0,1]^m \to \mathbb{R}$, each $i = 1, \ldots, m$, and each $y \in [0,1]^{m-1}$ such that the function $w(\rho(\cdot, i, y))$ has bounded variation on $[0,1]$, define $\Phi(w, i, y)$ to be the total variation on $[0,1]$ of $w(\rho(\cdot, i, y))$.

Theorem 4.2: Let $b, \gamma, \bar{\Phi} > 0$. Define $S_2(b, \gamma, \bar{\Phi})$ as the set of all $s \in L^{\ell}_{\infty}(\mathbb{R}^m)$ such that

(i) $\|s\|_\infty \leq b$, and

(ii) for every $x \in \mathbb{R}^m$ and every $i \in \{1, \ldots, m\}$, the total variation of each component of the map from $[0, \gamma]$ to $\mathbb{R}^\ell$ given by

$$s(x + \rho(\tau, i, 0_{m-1})), \qquad \tau \in [0, \gamma], \tag{4.3}$$

is no greater than $\bar{\Phi}$.

then $S_2(b, \gamma, \bar{\Phi})$ is myopically L_p–compact for every $p \geq 1$.

We discuss the significance of Theorem 4.2 before giving its proof. Observe first that the total variation of each component of the map (4.1) of Theorem 4.1 is bounded by $\bar{\Phi} = \gamma\lambda + 2b\kappa$, where λ and κ are as in Theorem 4.1. Therefore

$$S_1(b, \gamma, \lambda, \kappa) \subseteq S_2(b, \gamma, (\gamma\lambda + 2b\kappa)),$$

from which it follows easily that $S_1(b, \gamma, \lambda, \kappa)$ is relatively compact if $S_2(b, \gamma, (\gamma\lambda + 2b\kappa))$ is relatively compact. Therefore Theorem 4.1 follows immediately from Theorem 4.2.

Now consider the case in which a feature edge in an image is slightly blurred by poor focus, by limitations of a camera lens, or by imperfect corners of objects in the image. We cannot count such an edge as a discontinuity in Theorem 4.1. Instead we need to make λ very large to account for a very sudden change. However, the variation of a feature edge does not increase if the edge is blurred, as long as the blurred edge changes monotonically. So we can let κ' be the number of feature edge crossings, in which we may include not only true discontinuities, but also any monotonic changes in image amplitude we wish to include. Then the image will be an element of $S_2(b, \gamma, (\gamma\lambda + 2b\kappa'))$ under conditions similar to those of Theorem 4.1. This is helpful because in some cases we may be unable to tell whether a particular feature edge is truly a discontinuity, or only a very sudden monotonic change.

To expand on this idea, suppose we have a complicated image, such as the snapshot of Palmer Story in Figure 4.3.[10] In some portions of the image, such as Palmer's hair and eyebrows, feature edges occur so close together that image imperfections do not allow us to distinguish all the features. Such imperfections can be caused by, for example, poor focus, an imperfect lens, or sampling. For such a complicated image, we may choose λ to be anything we like, and then let κ' correspond to the maximum number of crossings of monotonic feature edges which exceed the rate of change in image intensity specified by λ.

4.4.2 Proof of Theorem 4.2

Here we will prove Theorem 4.2. First we will need the following proposition, which is proved in Appendix A.

Proposition 4.1: Let ℓ and m be positive integers, and let $U \subseteq L_p^\ell([0,1]^m)$. U is relatively compact if there are $b, \bar{\Phi} > 0$ such that for each $u \in U$,

(i) $\|u\|_\infty \leq b$, and

[10] Palmer Story is the author's nephew.

Figure 4.3: A snapshot of the author's nephew, Palmer Story

(ii) for every $k \in \{1, \ldots, \ell\}$, every $i \in \{1, \ldots, m\}$, and every $y \in [0,1]^{m-1}$, $u_k(\rho(\cdot, i, y))$ has bounded variation on $[0,1]$, and

$$\Phi(u_k, i, y) \leq \bar{\Phi}.$$

We need the following definition to prove Theorem 4.2. If $\varepsilon > 0$, X is a normed linear space with norm $\|\cdot\|$, and $E, M \subseteq X$, we say E is an "ε-net"[11] for M if for every $x \in M$ there is a $y \in E$ such that $\|x - y\| < \varepsilon$.

Proof of Theorem 4.2: Choose $p \geq 1$, and select $b, \gamma, \bar{\Phi} > 0$. Condition (i) of the theorem immediately satisfies condition (i) of the definition of a myopically L_p–compact set. Clearly $S_2(b, \gamma, \bar{\Phi})$ is also translation-invariant.

It remains to show that condition (ii) of the definition of a myopically L_p–compact set is satisfied. Let $r > 0$, and let α be the smallest integer such that $\gamma\alpha \geq r$. For each $s \in S_2(b, \gamma, \bar{\Phi})$, each integer $i \in \{1, \ldots, m\}$, each integer $j \in \{-\alpha, \ldots, \alpha - 1\}$, and each $y \in [-\gamma\alpha, \gamma\alpha]^{m-1}$, we can see by setting $x = \rho(j\gamma, i, y)$ in (4.3) that the total variation of each component of the map from $[j\gamma, (j+1)\gamma]$ to $\mathbb{R}^\ell$ given by

$$s(\rho(\tau, i, y)), \qquad \tau \in [j\gamma, (j+1)\gamma]$$

is no greater than $\bar{\Phi}$. It follows that the total variation of each component of the map from $[-\gamma\alpha, \gamma\alpha]$ to $\mathbb{R}^\ell$ given by

$$s(\rho(\tau, i, y)), \qquad \tau \in [-\gamma\alpha, \gamma\alpha],$$

is no greater than $2\alpha\bar{\Phi}$. Because dilation and translation do not change the total variation, we can say that for each $i \in \{1, \ldots, m\}$ and each $y \in [0,1]^{m-1}$, the total

[11] We use the definition of [29] §9, except that we state the definition in terms of normed linear spaces, since all the metric spaces in this chapter are normed linear spaces.

variation of each component of any map from $[0,1]$ to $\mathbb{R}^\ell$ given by

$$s(2\gamma\alpha\rho(\tau - \tfrac{1}{2}, i, y)), \qquad \tau \in [0,1],$$

where $s \in S_2(b, \gamma, \bar{\Phi})$, is no greater than $2\alpha\bar{\Phi}$. Therefore, with $w = s(2\gamma\alpha\,\cdot)$, we have that

$$\Phi(w_k, i, y) \leq 2\alpha\bar{\Phi}$$

for every $k \in \{1, \ldots, \ell\}$, every $i \in \{1, \ldots, m\}$, and every $y \in [0,1]^{m-1}$. Because dilation and translation do not change the topology of $L_p^\ell([0,1]^m)$ either, it follows from Proposition 4.1 that the set

$$\{s|_{[-\gamma\alpha,\gamma\alpha]^m} : s \in S_2(b, \gamma, \bar{\Phi})\} \tag{4.4}$$

of restrictions of elements of $S_2(b, \gamma, \bar{\Phi})$ to $[-\gamma\alpha, \gamma\alpha]^m$ is relatively compact, as a subset of $L_p^\ell([-\gamma\alpha, \gamma\alpha]^m)$.

Now by Theorem 3 of §9 of [29],[12] there exists a finite ε-net for the set (4.4). Note that $\bar{B}_r(0_m) \subseteq [-\gamma\alpha, \gamma\alpha]^m$. Restricting the elements of the ε-net for (4.4) to $\bar{B}_r(0_m)$ gives an ε-net for $\{s|_{\bar{B}_r(0_m)} : s \in S_2(b, \gamma, \bar{\Phi})\}$. Because $L_p^\ell([-\gamma\alpha, \gamma\alpha]^m)$ is complete, it follows from Theorem 3 of §9 of [29] that the set $\{s|_{\bar{B}_r(0_m)} : s \in S_2(b, \gamma, \bar{\Phi})\}$ of restrictions of elements of $S_2(b, \gamma, \bar{\Phi})$ to $\bar{B}_r(0_m)$ is relatively compact. Therefore $S_2(b, \gamma, \bar{\Phi})$ is myopically L_p–compact. This completes the proof.

[12]In [29], as in many Russian mathematical texts, "compact" means what we call relatively compact, and "compact in itself" means what we call compact.

4.5 Myopically L_p–Compact Sets and Arc Length of Feature Edges

4.5.1 Theorem and Discussion

In Theorem 4.3 below, we direct attention to the special case $m = 2$. Theorem 4.3 is similar to Theorem 4.1 in that it gives a myopically L_p–compact set of functions characterized by a magnitude bound, a condition regulating the location of discontinuities, and a Lipschitz condition on regions of continuity. However, in Theorem 4.3, the discontinuities are constrained to lie along rectifiable curves, and we place a bound on the number and length of the curves. We also use a different Lipschitz condition.

We begin with the following definitions. We say that a map $u : D \to \mathbb{R}^\ell$, where D is a subset of $\mathbb{R}^2$, is locally Lipschitz on an open $A \subseteq D$, with Lipschitz constant λ, if for every $x \in A$, there is an $r > 0$ such that the restriction of u to $B_r(x) \bigcap D$ is Lipschitz with constant λ, i.e.,

$$x', x'' \in B_r(x) \bigcap D \quad \Rightarrow \quad |u(x') - u(x'')|_1 \le \lambda |x' - x''|_2.$$

As an example of a locally Lipschitz map which is not Lipschitz, consider the following map $u : [-2, 2]^2 \to \mathbb{R}^2$:

$$u(x) = \begin{cases} 1 - x_2, & -1 \le x_1 \le 1, \quad 0 \le x_2 \le 1 \\ 1 + x_2, & -1 \le x_1 \le 1, \quad -1 \le x_2 < 0 \\ 0, & \text{otherwise} \end{cases} \qquad x \in \mathbb{R}^2. \tag{4.5}$$

See the illustration in Figure 4.4. Let $A' = \{-1, 1\} \times [-1, 1]$. Clearly u is not Lipschitz, because u is discontinuous on A'. Nor is u Lipschitz on $\mathbb{R}^2 - A'$, because if $0 < \delta < 1$, the two points $x = (1 + \delta, 0)$ and $x' = (1 - \delta, 0)$ may be arbitrarily close to one another as δ is made small, but $|u(x) - u(x')| = 1$. However, one can see that u is locally Lipschitz on $A = \mathbb{R}^2 - A'$, with Lipschitz constant 1.

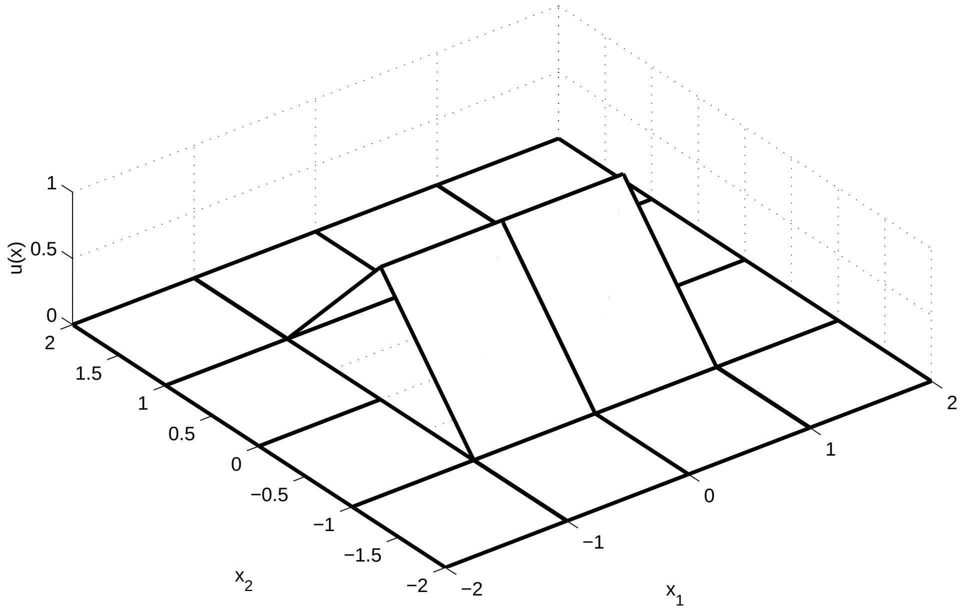

Figure 4.4: An example locally Lipschitz $u : [-2, 2]^2 \to \mathbb{R}$

Let $a, b \in \mathbb{R}$ with $a < b$. Suppose $f : [a, b] \to \mathbb{R}^2$ continuous. We say f describes a rectifiable curve in $\mathbb{R}^2$ if $\{\Lambda(f, P) : P \in \mathcal{P}(a, b)\}$ has an upper bound, where $\Lambda(f, P)$ is defined by

$$\Lambda(f, P) = \sum_{i=1}^{k_P} |f(t_i) - f(t_{i-1})|_2,$$

and where $\{t_0, \ldots, t_{k_P}\}$ is the partition P. If f describes a rectifiable curve, we say the arc length of f is

$$\bar{\Lambda}(f) = \sup\{\Lambda(f, P) : P \in \mathcal{P}(a, b)\}.$$

If $t \in (a, b]$, then the restriction $f|_{[a,t]}$ of f to $[a, t]$ also describes a rectifiable curve (see Theorem 6.18 of [46]). We will denote its arc length by

$$\bar{\Lambda}(f, t) = \bar{\Lambda}(f|_{[a,t]}).$$

For convenience, we will define $\bar{\Lambda}(f, 0) = 0$.

Theorem 4.3: Let $b, \gamma, \lambda, \beta$ be real numbers such that $b, \gamma > 0$ and $\lambda, \beta \geq 0$, and let κ be a positive integer. Define $S_3(b, \gamma, \lambda, \beta, \kappa)$ to be the set of all $s \in L^{\ell}_{\infty}(\mathbb{R}^2)$ such that the following four conditions hold:

(i) $\|s\|_{\infty} \leq b$.

(ii) There is a closed set $A_s \subseteq \mathbb{R}^2$ such that s is locally Lipschitz, with Lipschitz constant λ, on $\mathbb{R}^2 - A_s$.

(iii) For every $x \in \mathbb{R}^2$, the set

$$\left\{x + y \in A_s : y \in [-\gamma, \gamma]^2\right\}$$

consists of points lying on no more than κ rectifiable curves, described by continuous functions $f_1^x, \ldots, f_\kappa^x : [0, 1] \to [0, 1]^2$.

(iv) For every $x \in \mathbb{R}^2$, the sum of the arc lengths of $f_1^x, \ldots, f_\kappa^x$ is no more than β, i.e.,

$$\sum_{i=1}^{\kappa} \bar{\Lambda}(f_i^x) \leq \beta.$$

Then $S_3(b, \gamma, \lambda, \beta, \kappa)$ is myopically L_p–compact for every $p \geq 1$.

We discuss the significance of Theorem 4.3 before proving it. Consider again the image shown in Figure 4.1. The image belongs to $S_3(b, \gamma, \lambda, \beta, \kappa)$ if we choose b, γ, λ, β, and κ as follows. We may choose b as we did in the discussion of Theorem 4.1. The value of λ is chosen similarly, except that we consider the rate of change in the image intensity in every direction, rather than only in the horizontal and vertical directions. The largest rate of change in image intensity occurs in the shadow beneath the near wing of the crane. So we may choose any λ which is no less than the maximum rate of change in the image intensity on a line segment crossing this shadow and perpendicular to it.

We then choose a set A_s consisting of rectifiable curves outlining all feature edges. A_s is pictured in Figure 4.5. We fix $\gamma > 0$, and look for an upper bound, κ, on the number of these rectifiable curves in a square of the form $x + [-\gamma, \gamma]^2$. We also find an upper bound, β, on the total length of the portions of these curves inside a square of the form $x + [-\gamma, \gamma]^2$. Note that the two squares yielding the maximum values of κ and β do not necessarily have the same center x. However, in this case, for the size square we have chosen, the two squares are the same. In the square drawn in the figure, we count 8 rectifiable curves having a total arc length of about 18γ. (Observe that γ is half the length of a side of the square.) Therefore we must have $\kappa \geq 8$ and $\beta \geq 18\gamma$. Note that a rectifiable curve can have a corner. Also note that if we wish, we may include points in A_s which are not points of discontinuity, but which allow us to join curves together which do outline discontinuities. This decreases the number of the curves, but increases the total arc length of the joined curves.

As additional support for the significance of Theorem 4.3, consider a digital radar or sonar image. As we discussed earlier, such an image may be represented by a map from $\mathbb{R}^2$ to $\mathbb{R}$ consisting of patches separated by arcs centered at the transducer, and by radial lines emanating from the transducer. The value of the map within each patch is determined by the digital sample in the center of the patch. Discontinuities of such a map occur only along patch edges. For a radar or sonar image with 100 range samples and 40 beams, with samples separated in range by 1 and in angle by $3°$, the edges are pictured in Figure 4.6. These edges

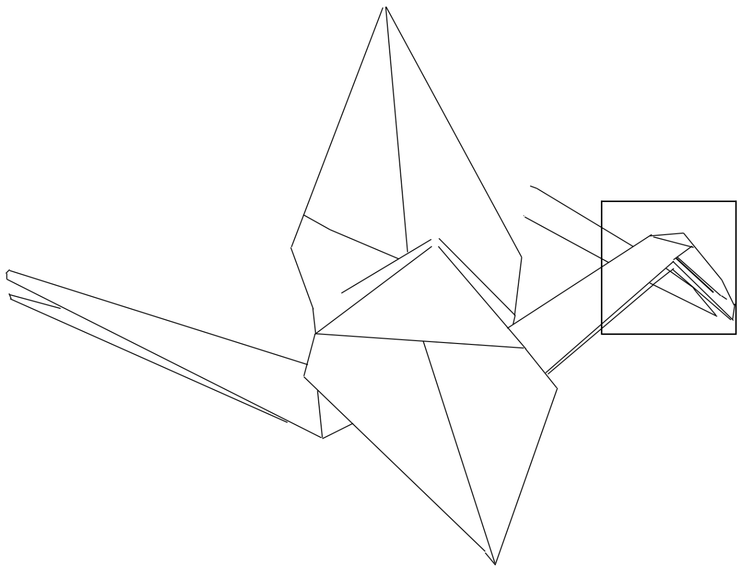

Figure 4.5: Set A_s of feature edges for Figure 4.1

form the set A_s. For such a radar or sonar system with a bounded dynamic range, every image produced by the system must belong to $S_3(b, \gamma, \lambda, \beta, \kappa)$ if we choose b, γ, λ, β, and κ as follows. Set b to the maximum output value of the radar or sonar system. Let $\lambda = 0$, since the value of the image can only change at patch edges. Choose any $\gamma > 0$. The portion of an arc that may be contained in a square of the form $x + [-\gamma, \gamma]^2$ may have arc length of at most $\frac{4\pi}{3\sqrt{3}}\gamma$.[13] No more than κ_a arcs may intersect such a square, where κ_a is the smallest integer greater than $2\gamma\sqrt{2}$. Further, any arc which intersects a square in more than one place may be replaced by a single, shorter curve which traces the edge of the square to connect the parts of the arc inside the square, so that the new curve does not leave the square. Therefore there are no more than κ_a curves created by the arcs within the square, and the total arc length is bounded by $\frac{4\pi}{3\sqrt{3}}\gamma\kappa_a$. The portion of a radial line that may be contained within such a square may be no longer than the diagonal of the square, $2\gamma\sqrt{2}$. All 41 radial lines (i.e. beam edges) intersect any square which contains the transducer, because all the radial lines emanate from that point. Therefore we may let $\kappa = \kappa_a + 41$, and $\beta = \frac{4\pi}{3\sqrt{3}}\gamma\kappa_a + 82\gamma\sqrt{2}$.

4.5.2 Proof of Theorem 4.3

We will need a few propositions to prove Theorem 4.3. Propositions 4.2 and 4.3 below are proved in Appendix A.

Proposition 4.2: Let $f : [0, 1] \to \mathbb{R}^2$ be a continuous function representing a rectifiable curve. Let $r > 0$, and let $\Theta(f, r)$ be the set of $x \in \mathbb{R}^2$ that are within r of some point on the curve, i.e.,

$$\Theta(f, r) = \{x \in \mathbb{R}^2 : |f(t) - x|_2 < r \quad \text{for some } t \in [0, 1]\}.$$

[13]This value is for an arc stretching across the entire square, tracing out the full 120° sector (40 beams × 3°) within the square. If an arc traces out less than the full sector within the square, then the portion of the arc within the square must have an arc length smaller than this.

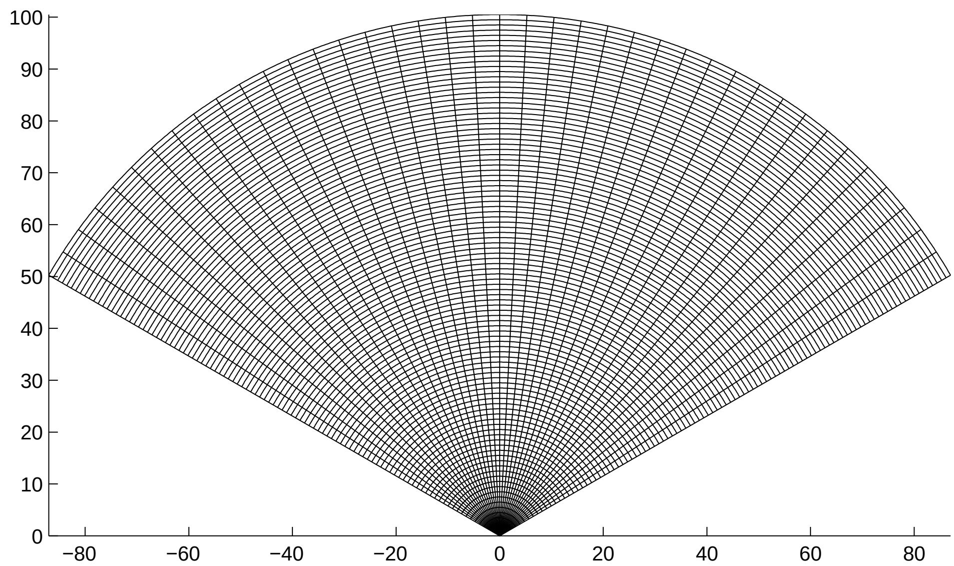

Figure 4.6: Set A_s for a radar or sonar image

Then $\mu(\Theta(f,r)) \le \pi r^2 + 2r\bar{\Lambda}(f)$, where μ denotes Lebesgue measure (i.e. area).

Proposition 4.3: Suppose $u \in L_\infty^\ell([0,1]^2)$ with $\|u\|_\infty \le b$. Assume $A \subseteq [0,1]^2$ is closed and $\lambda > 0$, and let

$$F = \left\{ x \in [0,1]^2 : \inf_{y \in A} |x-y|_2 \ge \frac{\ell b}{\lambda} \right\}.$$

If u is locally Lipschitz on $A' = [0,1]^2 - A$ with Lipschitz constant λ, then the restriction $u|_F$ of u to F is Lipschitz with Lipschitz constant λ.

The following proposition is the key to the proof of Theorem 4.3. We include the proof of this proposition here because the most important elements of the proof of Theorem 4.3 occur within the proof of Proposition 4.4.

Proposition 4.4: Let $b, \lambda, \beta \in \mathbb{R}$ with $b > 0$ and $\lambda, \beta \ge 0$, and let κ be a positive integer. Let U be a subset of $L_\infty^\ell([0,1]^2)$ such that the following four conditions hold for every $u \in U$:

(i) $\|u\|_\infty \le b$.

(ii) There is a closed set $A_u \subseteq [0,1]^2$ such that u is locally Lipschitz, with Lipschitz constant λ, on $[0,1]^2 - A_u$.

(iii) A_u consists of the points lying on no more than κ rectifiable curves, described by continuous functions $f_1, \ldots, f_\kappa : [0,1] \to [0,1]^2$.

(iv) The sum of the arc lengths of $f_1, \ldots, f_\kappa$ is no more than β, i.e.,

$$\sum_{i=1}^{\kappa} \bar{\Lambda}(f_i) \le \beta.$$

Then U is relatively compact in $L_p^\ell([0,1]^2)$ for every $p \ge 1$.

Proof: Fix $p \geq 1$, and choose $\varepsilon > 0$. Let $\lambda' > 0$ satisfy

$$\lambda' \geq \lambda \qquad \text{and} \qquad 2b\left(\kappa\pi\left(\frac{\ell b}{\lambda'}\right)^2 + 2\beta\frac{\ell b}{\lambda'}\right)^{\frac{1}{p}} < \varepsilon \tag{4.6}$$

Since $\lambda' \geq \lambda$, each $u \in U$ must locally Lipschitz with Lipschitz constant λ' on $[0,1]^2 - A_u$.

For each $u \in U$, let

$$F_u = \left\{x \in [0,1]^2 : \inf_{y \in A_u} |x-y|_2 \geq \frac{\ell b}{\lambda'}\right\}.$$

By Proposition 4.3, the restriction $u|_{F_u}$ of each u to F_u is Lipschitz, with Lipschitz constant λ'. Each F_u is closed, because F_u is the complement in $[0,1]^2$ of the union of all open balls of radius $\frac{\ell b}{\lambda'}$ centered in A_u. Denote the restriction of u to F_u by $u|_{F_u}$. It is shown in [47] that there is a continuous extension v_u of $u|_{F_u}$ to all of $[0,1]^2$ such that v_u is Lipschitz on $[0,1]^2$ with Lipschitz constant λ', and such that $\|v_u\|_\infty \leq b$. Because this bound and this Lipschitz condition hold for every v_u such that $u \in U$, it follows from the Arzela-Ascoli theorem (see Theorem IV.6.7 of [48], for example) that $\{v_u : u \in U\}$ is relatively compact in $L^\ell_\infty([0,1]^m)$. But for $v \in L^\ell_\infty([0,1]^m)$, we have $\|v\|_p \leq \ell\|v\|_\infty$, and therefore $\{v_u : u \in U\}$ is also relatively compact in $L^\ell_p([0,1]^m)$.

We can see from (iii) that in the notation of Proposition 4.2,

$$F_u = \bigcup_{i=1}^{\kappa} \Theta\left(f_i, \frac{\ell b}{\lambda'}\right).$$

So by (iv) and Proposition 4.2,

$$\mu(F_u) \leq \kappa\pi\left(\frac{\ell b}{\lambda'}\right)^2 + 2\beta\frac{\ell b}{\lambda'}.$$

Since $u = v_u$ except on F_u we have

$$\|v_u - u\|_p = \left(\int_{F_u} |v_u(x) - u(x)|_1^p dx \right)^{\frac{1}{p}}.$$

But since $\|u\|_\infty \leq b$ and $\|v_u\|_\infty \leq b$,

$$\|v_u - u\|_p \leq 2b\big(\mu(F_u)\big)^{\frac{1}{p}} \leq 2b\left(\kappa\pi \left(\frac{\ell b}{\lambda'} \right)^2 + 2\beta \frac{\ell b}{\lambda'} \right)^{\frac{1}{p}} < \varepsilon,$$

using (4.6). So $\{v_u : u \in U\}$ is a relatively compact ε-net for U. Therefore by Corollary I of §I.9 of [29], U is relatively compact. This completes the proof.

We are now ready to prove Theorem 4.3.

Proof of Theorem 4.3: Choose $p \geq 1$ and $r > 0$. Clearly $S_3(b, \gamma, \lambda, \beta, \kappa)$ is shift-invariant. We need to show that the set of restrictions of elements of $S_3(b, \gamma, \lambda, \beta, \kappa)$ to $\bar{B}_r(0_2)$ is relatively compact.

Let α be the smallest integer such that $\gamma\alpha \geq r$. For $i = 0, \ldots, \alpha - 1$, let I_i be the interval

$$I_i = \big[(2i - \alpha)\gamma,\ (2(i+1) - \alpha)\gamma\big].$$

For integers i and j between 0 and $\alpha - 1$, inclusive, let $Q_{i,j}$ be the square given by $Q_{i,j} = I_i \times I_j$. Note that

$$[-\alpha\gamma, \alpha\gamma]^2 = \bigcup_{i=0}^{\alpha-1} \bigcup_{j=0}^{\alpha-1} Q_{i,j}.$$

Let U be the set of restrictions of elements of $S_3(b, \gamma, \lambda, \beta, \kappa)$ to $[-\alpha\gamma, \alpha\gamma]^2$. Fix $u \in U$, and let s be an element of $S_3(b, \gamma, \lambda, \beta, \kappa)$ such that u is the restriction of s to $[-\alpha\gamma, \alpha\gamma]^2$. Set $A_u = A_s \bigcap [-\alpha\gamma, \alpha\gamma]^2$. Clearly $\|u\|_\infty \leq b$, and clearly u is locally Lipschitz, with Lipschitz constant λ, on $[-\alpha\gamma, \alpha\gamma]^2 - A_u$. Observe that

$$A_u = \bigcup_{i=0}^{\alpha-1} \bigcup_{j=0}^{\alpha-1} \Big(Q_{i,j} \bigcap A_s \Big).$$

But by (iii) and (iv), each $Q_{i,j} \bigcap A_s$ consists of points lying on no more than κ rectifiable curves whose arc lengths sum to no more than β. Because A_u contains α^2 such regions, the points of A_u lie on no more than $\kappa' = \alpha^2\kappa$ rectifiable curves whose arc lengths sum to no more than $\alpha^2\beta$. This holds true for every $u \in U$, so by Proposition 4.4, U is relatively compact in $L_p^\ell([-\alpha\gamma, \alpha\gamma]^2)$.

By Theorem 3 of §9 of [29], there exists a finite ε-net for U. Note that $\bar{B}_r(0_2) \subseteq [-\gamma\alpha, \gamma\alpha]^2$. Restricting the elements of the ε-net for U to $\bar{B}_r(0_2)$ gives an ε-net for $\{s|_{\bar{B}_r(0_2)} : s \in S_3(b, \gamma, \lambda, \beta, \kappa)\}$. Because $L_p^\ell([-\gamma\alpha, \gamma\alpha]^2)$ is complete, it follows from Theorem 3 of §9 of [29] that the set $\{s|_{\bar{B}_r(0_2)} : s \in S_3(b, \gamma, \lambda, \beta, \kappa)\}$ of restrictions of elements of $S_3(b, \gamma, \lambda, \beta, \kappa)$ to $\bar{B}_r(0_2)$ is relatively compact, and so $S_3(b, \gamma, \lambda, \beta, \kappa)$ is myopically L_p–compact. This completes the proof.

Chapter 5

L-Myopic Systems

5.1 Introduction

As we discussed in Chapter 1, dynamic systems play an important role in engineering and science, and nonlinear dynamic systems are often difficult to analyze directly. It is often helpful to represent such a system, or an approximation to such a system, by a combination of familiar elements which are easier to analyze. One such approach is to use a Volterra series, or a Volterra series-like expansion, to represent or to approximate a system using a sum of iterated integrals (see [9] and [6], as well as papers referenced there). Another approach is to approximate a system by a bank of linear dynamic systems, followed by a nonlinear memoryless system ([10], [44], [14], [12]). (See [13] for a discussion of each approach, and a more complete bibliography.[1]) In both approaches, whenever the approximation is to hold for all time, a central condition is that the value of the output of a system at a given time must be relatively independent of the values of the input at remote times.[2] The terms "fading memory" ([6], [14], [50], [51]), "decaying memory" ([52]), "approximately finite

[1]For a collection of recent results that are related in a general sense, see [18].

[2]An exception to this appears in [49], in which a system without this property may be tracked arbitrarily closely over all time. However, there the approximating structure has access to the outputs of the system as well as its inputs.

memory" ([10]), and "myopic" ([44], [12]) are among the names for different ways of making this general idea precise, in different settings. It should be noted that the terms are not always good indications of the underlying definitions. "Fading memory" in [50] and in [51], for example, is more similar to "approximately finite memory" in [26] than to "fading memory" in [6] and [49], and although the meanings of "myopic" in [44] and [12] are closely related, the differences are important.

The general idea behind fading memory and the other similar concepts is intuitively reasonable. However, unless it is shown that some familiar class of systems actually has one of these properties, an engineer or scientist may doubt whether the results derived for such systems have any real use. This question is addressed in [11] and [53], in which it was shown that a feedback system satisfying the familiar circle criterion must have approximately finite memory, and therefore may be approximated. Later, these results were used in [54] to show that a feedback system with approximately finite memory may be viewed as a map that has fading memory in the sense of [6], and that is myopic in the sense of [44]. This establishes the applicability of the approximation results in those papers also.

One of the difficulties addressed in [54] is that the definition of a map with fading memory and the definition of a myopic map assume that the maps have inputs and outputs defined on all of $\mathbb{R}$. For causal systems, this is a more abstract setting than that of [11] and [53], in which inputs and outputs are defined on $\mathbb{R}_+$, the nonnegative real numbers. Because of this, it was necessary to extend[3] the feedback system map to a map having inputs and outputs defined on all of $\mathbb{R}$. Similar results appeared earlier in [52] and [26], in the context of characterizing steady-state responses to almost-periodic inputs.

[3]Here our use of the terms "extend" and "extension" differs from what is usual. Normally, to "extend" a map means to enlarge the domain of the map. Technically, the domain of the map here is not enlarged; rather it is changed entirely. However, the domain of the original map is closely related to a subset of the domain of the "extended" map, so that the resulting map actually is an extension, in the usual sense, of a closely related map. See Section 5.3.1 and Theorem 5.1 for a precise statement of what we mean.

The definition of a myopic system in [12] is attractive because while other approximation results for systems with inputs and outputs defined on $\mathbb{R}$ ([6], [44]) require that the inputs be continuous, the results in [12] allow the approximation of continuous-time systems with inputs that may have discontinuities. However, no familiar class of systems has yet been shown to be myopic in the sense of [12].

The purpose of this chapter is to show that a certain familiar system is myopic in the sense of [12]. This establishes the relevance of the approximation results in [12]. Specifically, it is shown here that if an input-output map represents a feedback system satisfying the circle criterion, then a certain important extension[3] of that map, to a map with inputs and outputs defined on $\mathbb{R}$, is myopic in the sense of [12].

This chapter is organized as follows. In Section 5.2, we establish the notation used throughout the chapter, and describe the feedback system G_N. In Section 5.3, we make minor modifications to a theorem in [26]. This allows us to extend[3] G_N to a map F_N with inputs and outputs defined on $\mathbb{R}$. Also, a continuity condition needed in Section 5.4 is shown to hold for F_N. Section 5.4 presents a theorem which shows that F_N is myopic. Further, it is observed that a consequence of the theorem is that two parameters used to define a myopic system are unnecessary. That is, a system that is myopic with respect to one $p \geq 1$ and one weighting function w (with the required properties) is myopic with respect to every $p \geq 1$ and every weighting function w (again, with the required properties). Appendix B, which is related to this chapter, gives a necessary and sufficient condition for a system to have approximately finite memory. This condition has appeared in [26] and elsewhere as a condition "related" to the definition of approximately finite memory, but it was not previously observed to be equivalent.

5.2 Preliminaries

5.2.1 General Preliminaries

The following definitions are used throughout this chapter. Let $\mathbb{R}$ and $\mathbb{R}_+$ be the set of real numbers and of nonnegative real numbers, respectively. For some positive integer n, let $\mathbb{R}^n$ be the set of real n-vectors, with zero element 0_n, and with the Euclidean norm $|\cdot|$. (Note that we also use $|\cdot|$ to denote the absolute value of a real number, but the context makes clear what is meant.)

For $1 \leq p < \infty$, and with D either $\mathbb{R}$, $\mathbb{R}_+$, or a subinterval of $\mathbb{R}$, let $L_p(D)$ be the set of Lebesgue integrable maps $v : D \to \mathbb{R}^n$ with $\int_D |v(\tau)|^p \, d\tau < \infty$. Define the usual norm on $L_p(D)$ by $\|v\|_p = \left(\int_D |v(\tau)|^p \, d\tau\right)^{\frac{1}{p}}$. (Actually, we are using the notation $\|\cdot\|_p$ to denote the norm on many different normed spaces — one for each choice of D. However, D is clear from the context, and this should cause no confusion.) Let $L_\infty(D)$ be the set of essentially bounded Lebesgue measurable v from D to $\mathbb{R}^n$. Let $L_{\infty e}(D)$ be the set of Lebesgue measurable functions $v : D \to \mathbb{R}^n$ that are bounded on bounded subintervals of D. Define the norm $\|\cdot\|_\infty$ on $L_\infty(D)$ by

$$\|v\|_\infty = \operatorname{ess\,sup}_{\tau \in D} \left(\max_{1 \leq i \leq n} |v_i(\tau)| \right),$$

where v_i is the i^{th} component of v.[4]

Let b be a positive number, and let S be the set of elements $s \in L_\infty(\mathbb{R})$ with $|s(t)| \leq b$ for all $t \in \mathbb{R}$. Similarly, let S_+ be defined as the set of elements $u \in L_\infty(\mathbb{R}_+)$ with $|u(t)| \leq b$ for all $t \in \mathbb{R}_+$. Let V and V_+ be the sets of all functions of the form $s : \mathbb{R} \to \mathbb{R}^n$ and $u : \mathbb{R}_+ \to \mathbb{R}^n$, respectively.

For each $\alpha \geq 0$, define the time delay map $T_\alpha : V_+ \to V_+$ by $(T_\alpha u)(t) =$

[4] In this chapter, as is customary, we will not explicitly differentiate between $L_p(D)$ viewed as a set of functions, and $L_p(D)$ viewed as a metric space. Functions which differ only on a set of Lebesgue measure zero are considered to be the same element of the metric space. Likewise, we will not explicitly differentiate between $L_\infty(D)$ or $L_{\infty e}(D)$ viewed as a set of functions and $L_\infty(D)$ or $L_{\infty e}(D)$ (respectively) viewed as a metric space.

$u(t-\alpha)$ for $t \geq \alpha$ and $(T_\alpha u)(t) = 0_n$ for $0 \leq t < \alpha$. For each $\alpha \in \mathbb{R}$, define another time delay map $\tilde{T}_\alpha : V \to V$ by $(\tilde{T}_\alpha s)(t) = s(t-\alpha)$ for every $t \in \mathbb{R}$. Note that S and S_+ are shift-invariant, i.e., $T_\alpha S_+ \subseteq S_+$ for every $\alpha \geq 0$, and $\tilde{T}_\alpha S \subseteq S$ for every $\alpha \in \mathbb{R}$. We say $G : S_+ \to V_+$ is time-invariant if for every $u \in S_+$ and every $\alpha \geq 0$, $T_\alpha G u = G T_\alpha u$. Similarly, we say $F : S \to V$ is time-invariant if for every $s \in S$ and every $\alpha \in \mathbb{R}$, $\tilde{T}_\alpha F s = F \tilde{T}_\alpha s$.

We say $G : S_+ \to V_+$ is causal if for each $t \in \mathbb{R}_+$, and for every $u_1, u_2 \in S_+$ with $u_1(\alpha) = u_2(\alpha)$ for $0 \leq \alpha \leq t$, we have $(Gu_1)(t) = (Gu_2)(t)$. Similarly, we say $F : S \to V$ is causal if for each $t \in \mathbb{R}$, and for every $s_1, s_2 \in S$ with $s_1(\alpha) = s_2(\alpha)$ for $\alpha \leq t$, we have $(Fs_1)(t) = (Fs_2)(t)$.

Define an extension map $E : S_+ \to S$ by $(Eu)(t) = u(t)$ for $t \geq 0$, and $(Eu)(t) = 0_n$ otherwise.[5] Define a projection map $P : S \to S_+$ by $(Ps)(t) = s(t)$ for all $t \in \mathbb{R}_+$. For each $a \in \mathbb{R}$, define $Q_a : S \to S$ by $(Q_a s)(t) = s(t)$ for $t \geq a$ and $(Q_a s)(t) = 0_n$ otherwise.

For each $\alpha, \beta > 0$, define a windowing function $W_{\beta,\alpha} : S_+ \to S_+$ by

$$(W_{\beta,\alpha} u)(t) = \begin{cases} u(t), & t \in [\beta - \alpha, \beta] \\ 0_n, & \text{otherwise} \end{cases}, \qquad t \in \mathbb{R}_+.$$

Define a similar windowing function $\tilde{W}_{\beta,\alpha} : S \to S$ for each $\alpha, \beta > 0$ by

$$(\tilde{W}_{\beta,\alpha} s)(t) = \begin{cases} s(t), & t \in [\beta - \alpha, \beta] \\ 0_n, & \text{otherwise} \end{cases}, \qquad t \in \mathbb{R}.$$

We say $G : S_+ \to V_+$ has approximately finite memory if for every $\varepsilon > 0$,

[5] Here we use "extension" in the usual sense, in that the domain of u is enlarged.

there is a $\Delta \geq 0$ such that

$$|(Gu)(t) - (GW_{t,\alpha}u)(t)| < \varepsilon$$

for all $u \in S_+$, all $t \in \mathbb{R}_+$ and all $\alpha \geq \Delta$. Similarly, we say $F : S \to V$ has approximately finite memory if for every $\varepsilon > 0$, there is a $\Delta \geq 0$ such that

$$|(Fs)(t) - (F\tilde{W}_{t,\alpha}s)(t)| < \varepsilon$$

for all $s \in S$, all $t \in \mathbb{R}$ and all $\alpha \geq \Delta$. Note that in either case, a map with approximately finite memory must be causal, because $\tilde{W}_{t,\alpha}s$ and $W_{t,\alpha}u$ are zero at times greater than t.

For any $p \geq 1$, we say $F : S \to V$ has continuity property $\mathcal{C}_p$ if for every $t \in \mathbb{R}$, the functional $(F\cdot)(t)$ is uniformly continuous under $\|\cdot\|_p$. By this we mean that for every $t \in \mathbb{R}$ and every $\varepsilon > 0$, there is a $\delta > 0$ such that if $s, s' \in S \cap L_p(\mathbb{R})$ and $\|s - s'\|_p < \delta$, then $|(Fs)(t) - (Fs')(t)| < \varepsilon$.

For each $\Delta > 0$, let $S_\Delta = \{u|_{[0,\Delta]} : u \in S_+\}$, where $u|_{[0,\Delta]}$ is the restriction of u to $[0, \Delta]$. If $G : S_+ \to V_+$ is causal and $\Delta > 0$, define the functional $G_\Delta : S_\Delta \to \mathbb{R}$ by

$$G_\Delta v = (Gu)(\Delta), \quad v \in S_\Delta, \tag{5.1}$$

where u is any element of S_+ such that $v = u|_{[0,\Delta]}$. For any $p \geq 1$, we say a causal $G : S_+ \to V_+$ has continuity property $\mathcal{C}_{p+}$ if for every $\Delta \in \mathbb{R}_+$, the functional G_Δ is uniformly continuous over S_Δ with respect to $\|\cdot\|_p$. By this we mean that for every $t \in \mathbb{R}_+$ and every $\varepsilon > 0$, there is a $\delta > 0$ such that if $v, v' \in S_\Delta$ and $\|v - v'\|_p < \delta$, then $|G_\Delta v - G_\Delta v'| < \varepsilon$.

5.2.2 A Familiar Class of Nonlinear Feedback Systems.

Here we return to the class of feedback systems considered in Section 3.6. It is shown in [53] (see also [11], [21]) that this familiar class of nonlinear feedback systems has approximately finite memory, and has continuity property $\mathcal{C}_{2+}$. The system is as follows (see Figure 3.1). The input r is an element of $L_{\infty e}(\mathbb{R}_+)$, and it is assumed that there is a solution such that e, w, and y also belong to $L_{\infty e}(\mathbb{R}_+)$. (This is a standard assumption, and it is typically satisfied.) Let r, e, w, and y be related by

$$\left.\begin{aligned} e(t) &= r(t) - y(t) \\ w(t) &= (\hat{\eta} e)(t) \\ y(t) &= (\mathcal{K} w)(t) + k_0(t) \end{aligned}\right. \qquad , t \geq 0 \tag{5.2}$$

where $k_0 \in L_\infty(\mathbb{R}_+)$ accounts for initial conditions, and where $\mathcal{K}$ and $\hat{\eta}$ are as follows. The operator $\mathcal{K}$ is defined by

$$(\mathcal{K}x)(t) = \int_0^t k(t-\tau)x(\tau)d\tau, \quad t \geq 0$$

for $x \in L_{\infty e}(\mathbb{R}_+)$. We assume k is a map from $\mathbb{R}_+$ to the real $n \times n$ matrices such that for each k_j, where k_j is the j^{th} column of k, the map $\xi_j : \mathbb{R}_+ \to \mathbb{R}^n$ given by $\xi_j(t) = t^p k_j(t)$ is in $L_1(\mathbb{R}_+) \cap L_2(\mathbb{R}_+)$ for $p \in \{0, 1, 2\}$. We further assume $\hat{\eta} : L_{\infty e}(\mathbb{R}_+) \to L_{\infty e}(\mathbb{R}_+)$ is a memoryless nonlinear map given by $(\hat{\eta}x)(t) = [\eta_1(x_1(t)), \ldots, \eta_n(x_n(t))]^T$ (where the superscript ^{T} denotes the matrix transpose) for $t \in \mathbb{R}_+$ and $x \in L_{\infty e}(\mathbb{R}_+)$. Assume that each η_j satisfies $\eta_j(0) = 0$ and $\alpha_0 \leq [\eta_j(a) - \eta_j(b)](a-b)^{-1} \leq \beta_0$ for real $a \neq b$, for some $\alpha_0, \beta_0 \in \mathbb{R}$ with $\alpha_0 \leq \beta_0$.

Assume further that:

$$\det[1_n + \frac{1}{2}(\alpha_0 + \beta_0)K(s)] \neq 0, \quad \text{Re}(s) \geq 0$$

and

$$\frac{1}{2}(\beta_0 - \alpha_0) \sup_{\omega \in \mathbb{R}} \Lambda\{[1_n + \frac{1}{2}(\alpha_0 + \beta_0)K(i\omega)]^{-1} K(i\omega)\} < 1,$$

in which 1_n is the identity matrix of order n, K is the Laplace transform of k, and $\Lambda\{\cdot\}$ denotes the largest singular value of $\{\cdot\}$.

When $n = 1$, the conditions above are satisfied if the circle criterion is met, i.e., if one of the following three conditions is satisfied:

1) $0 < \alpha_0 < \beta_0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ is bounded away from C_1 and does not encircle C_1, where C_1 is the circle of radius $\frac{1}{2}(\alpha_0^{-1} - \beta_0^{-1})$ centered on the real axis of the complex plane at $[-\frac{1}{2}(\alpha_0^{-1} + \beta_0^{-1}), 0]$.

2) $0 = \alpha_0 < \beta_0$, and $\mathrm{Re}[K(i\omega)] > -\beta_0^{-1}$ for all real ω.

3) $\alpha_0 < 0 < \beta_0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ is contained within the circle C_2 of radius $\frac{1}{2}(\beta_0^{-1} - \alpha_0^{-1})$ centered on the real axis of the complex plane at $[-\frac{1}{2}(\alpha_0^{-1} + \beta_0^{-1}), 0]$.

Under these conditions, it is shown in Section 2.3 of [53] that when $k_0(t) = 0_n$ for all $t \in \mathbb{R}_+$, there is a unique map $G_N : S_+ \to L_\infty(\mathbb{R}_+)$ taking r to y in accordance with (5.2), and that G_N has approximately finite memory. Furthermore, G_N has continuity property $\mathcal{C}_{2+}$ ([11], see also the comments in Section 2.4 of [53]), and is time-invariant.

It should be noted that G_N describes the system even in the case of nonzero initial conditions. Specifically, if the condition $k_0(t) = 0_n$ is not met, the relationship between the input r and the output y still depends on G_N according to the following equation:

$$y(t) = (G_N[r - k_0])(t) + k_0(t).$$

So, results shown to hold for G_N are not limited in scope to the case of zero initial conditions.

5.3 Extension of G_N to a map $F_N : S \to V$.

5.3.1 Existence and properties of F_N.

In Section 5.4, which follows, The term L_p-myopic refers to systems whose input and output sets must be S and V, respectively. But G_N takes S_+ to $L_\infty(\mathbb{R}_+)$. The following theorem allows us to extend[6] G_N to a map $F_N : S \to L_\infty(\mathbb{R})$. Since $L_\infty(\mathbb{R}) \subseteq V$, we then have $F_N : S \to V$. This theorem is a modification of Theorem 5 of [26]; see the proof of Proposition 5.1 below.

Theorem 5.1: Suppose $G : S_+ \to L_\infty(\mathbb{R}_+)$ is time-invariant and has approximately finite memory. Then there is an $F : S \to L_\infty(\mathbb{R})$ with the following properties:

(i) $(Fs)(t) = \lim_{a\to-\infty}(T_a GP\tilde{T}_{-a}s)(t)$ for every $s \in S, t \in \mathbb{R}$.

(ii) $(Gu)(t) = (FEu)(t)$ for every $u \in S_+, t \in \mathbb{R}_+$.

(iii) F is time-invariant.

(iv) $(Fs)(t) = \lim_{a\to-\infty}(FQ_a s)(t)$ for $s \in S$ and $t \in \mathbb{R}$.

(v) F has approximately finite memory, and is causal.

(vi) F is unique in the sense that if $H : S \to L_\infty(\mathbb{R})$ has properties (ii), (iii), and (iv), with F replaced with H, then $H = F$.

We need Propositions 5.1 and 5.2 below to prove the theorem.

In Proposition 5.1, we say F satisfies property P.1 if given $t_1 \in \mathbb{R}$ and $\varepsilon > 0$, there is a $t_2 < t_1$ such that

$$|(Fs)(t) - (FQ_{t_2}s)(t)| \leq \varepsilon, \quad t \geq t_1$$

[6] Technically, F_N is an extension not of G_N, but of a related map $G'_N : S'_+ \to L_\infty(\mathbb{R})$, where $S'_+ = \{Eu : u \in S_+\}$. This map is given for each $v \in S'_+$ by $(G'_N v)(t) = (G_N Pv)(t)$ for $t \geq 0$, and by $(G'_N v)(t) = 0$ for $t < 0$. See footnote 3.

for all $s \in S$.

Proposition 5.1: Under the conditions of Theorem 5.1, there is an $F : S \to L_\infty(\mathbb{R})$ satisfying (i), (ii), (iii), (iv), and (vi) of Theorem 5.1. Furthermore, F satisfies property P.1.

For the following proof, let P.2 denote the hypothesis that there is a nondecreasing $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\|Gu\|_\infty \leq \rho(\|u\|_\infty)$$

for every $u \in S_+$. Furthermore, we say G satisfies property P.3 if for each $s \in S$, q_s defined by

$$q_s(t) = \lim_{\alpha \to -\infty} (T_\alpha G P \tilde{T}_{-\alpha} s)(t), \quad t \in \mathbb{R}$$

satisfies $\|q_s\|_\infty \leq \rho(\|s\|_\infty)$.

Proof: In Theorem 5 of [26], it is shown that if the conditions of Theorem 5.1 are satisfied, and G also satisfies P.2, there is an $F : S \to L_\infty(\mathbb{R}_+)$ satisfying (i), (ii), (iii), and (iv), as well as properties P.1 and P.3. However, P.2 is only used to show that P.3 holds (see Theorem 2 and Section 2.2.3 of [52]). Specifically, under the conditions of our Theorem 5.1, (i), (ii), (iii), and (iv) are satisfied, and F has property P.1, even though P.2 is not satisfied. Finally, (vi) is shown to hold in Section 2.4.1 of [26], and again property P.2 is not used. This completes the proof of Proposition 5.1.

The following proposition is also used in the proof of Theorem 5.1; however, it is interesting in its own right as a sufficient (and necessary) condition for a system to have approximately finite memory. For this proposition we say that a time-invariant system $F : S \to V$ has property $\mathcal{A}_\tau$, for some $\tau \in \mathbb{R}$, if for every $\varepsilon > 0$, there exists a $\Delta \geq 0$ such that

$$|(Fs)(\tau) - (F\tilde{W}_{\tau,\Delta}s)(\tau)| < \varepsilon$$

for every $s \in S$.

Proposition 5.2: Suppose $F : S \to V$ is time-invariant, and suppose $\tau \in \mathbb{R}$. Then F has property $\mathcal{A}_\tau$ if and only if F has approximately finite memory.

It is important to note that this proposition does not hold if $F : S \to V$ is replaced with $G : S_+ \to V_+$ in the proposition and in the definition of $\mathcal{A}_\tau$. For an interesting similar proposition that holds under such conditions, see Appendix B.

Proof: Suppose F has approximately finite memory. Then taking $\alpha = \Delta$ and $t = \tau$, F has property $\mathcal{A}_\tau$. Conversely, suppose that F has property $\mathcal{A}_\tau$. Let $\varepsilon > 0$, and let Δ be the associate of $\frac{1}{2}\varepsilon$ in property $\mathcal{A}_\tau$. Let $s \in S_+$, $t \in \mathbb{R}$, and $\alpha > \Delta$. By the triangle inequality,

$$|(Fs)(t) - (F\tilde{W}_{\tau,\alpha}s)(t)| \leq |(Fs)(t) - (F\tilde{W}_{\tau,\Delta}s)(t)| + |(F\tilde{W}_{\tau,\Delta}s)(t) - (F\tilde{W}_{\tau,\alpha}s)(t)|. \tag{5.3}$$

Using the time-invariance of F, we have

$$(F\tilde{W}_{t,\alpha}s)(t) = (\tilde{T}_{\tau-t}F\tilde{W}_{t,\alpha}s)(\tau) = (F\tilde{W}_{\tau,\alpha}\tilde{T}_{\tau-t}s)(\tau).$$

Similarly, $(F\tilde{W}_{t,\Delta}s)(t) = (F\tilde{W}_{\tau,\Delta}\tilde{T}_{\tau-t}s)(\tau)$ and $(Fs)(t) = (F\tilde{T}_{\tau-t}s)(\tau)$. The right side of (5.3) becomes

$$|(F\tilde{T}_{\tau-t}s)(\tau) - (F\tilde{W}_{\tau,\Delta}\tilde{T}_{\tau-t}s)(\tau)| + |(F\tilde{W}_{\tau,\Delta}\tilde{T}_{\tau-t}s)(\tau) - (F\tilde{W}_{\tau,\alpha}\tilde{T}_{\tau-t}s)(\tau)|.$$

Property $\mathcal{A}_\tau$ bounds the first term on the right side of (5.3) by $\frac{1}{2}\varepsilon$, and because

$$F\tilde{W}_{\tau,\Delta}\tilde{T}_{\tau-t}s = F\tilde{W}_{\tau,\Delta}\big(\tilde{W}_{\tau,\alpha}\tilde{T}_{\tau-t}s\big),$$

property $\mathcal{A}_\tau$ also bounds the second term on the right side of (5.3) by $\frac{1}{2}\varepsilon$. So (5.3) gives

$$|(Fs)(t) - (F\tilde{W}_{\tau,\alpha}s)(t)| < \varepsilon,$$

which means that F has approximately finite memory. This completes the proof of Proposition 5.2.

Proof of Theorem 1: By Proposition 5.1, (i) – (iv) and (vi) hold, and F has property P.1. It remains only to show that (v) holds. We first show that F is causal. Let $t \in \mathbb{R}$, and let $s_1, s_2 \in S$ with $s_1(\tau) = s_2(\tau)$ for all $\tau \leq t$. Then by (i), the causality of G gives

$$(Fs_1)(t) = \lim_{\alpha\to-\infty}(T_\alpha GP\tilde{T}_{-\alpha}s_1)(t) = \lim_{\alpha\to-\infty}(T_\alpha GP\tilde{T}_{-\alpha}s_2)(t) = (Fs_2)(t), \quad t \in \mathbb{R}.$$

To show that F has approximately finite memory, we use Proposition 5.2. Let $t_1 \in \mathbb{R}$ and $\varepsilon > 0$. Let t_2 be the associate of $\frac{1}{2}\varepsilon$ and t_1 in property P.1, and set $t = t_1$. With $\Delta = t - t_2$, the causality of F gives us that $(FQ_{t_2}s)(t) = (FW_{t,\Delta}s)(t)$ for each $s \in S$. We have

$$|(Fs)(t) - (F\tilde{W}_{t,\Delta}s)(t)| = |(Fs)(t) - (FQ_{t_2}s)(t)| \leq \frac{1}{2}\varepsilon < \varepsilon, \quad t = t_1.$$

So F has property $\mathcal{A}_\tau$, and by Proposition 5.2, F has approximately finite memory. Therefore (v) holds, and the proof of Theorem 5.1 is complete.

Because we are interested in approximately finite memory, the following corollary serves as an interesting alternative to property (vi) of Theorem 5.1.

Corollary 5.1: Under the conditions of Theorem 5.1, F is unique in the sense that if a time-invariant $H : S \to L_\infty(\mathbb{R})$ has approximately finite memory and $(Gu)(t) = (HEu)(t)$ for $t \in \mathbb{R}_+$ and $s \in S_+$, then $H = F$.

Proof: Because of (vi) of Theorem 5.1, it suffices to show that for any H that has approximately finite memory, $(Hs)(t) = \lim_{a\to-\infty}(HQ_a s)(t)$ for every $s \in S$ and every $t \in \mathbb{R}$. Let $s \in S$, $t \in \mathbb{R}$, and $\varepsilon > 0$. Let Δ be the associate of $\frac{1}{2}\varepsilon$ regarding the approximately finite memory of H. Let $a < t - \Delta$. By the triangle inequality,

$$|(Hs)(t) - (HQ_a s)(t)| \leq |(Hs)(t) - (H\tilde{W}_{t,t-a}s)(t)| + |(H\tilde{W}_{t,t-a}s)(t) - (HQ_a s)(t)|.$$

The causality of H gives $(H\tilde{W}_{t,t-a}s)(t) = (H\tilde{W}_{t,t-a}Q_a s)(t)$. So approximately finite memory bounds both terms on the right side by $\frac{1}{2}\varepsilon$, and we have $|(Hs)(t) - (HQ_a s)(t)| < \varepsilon$. This completes the proof.

It is important to note that F is not the only time-invariant map such that $(Gu)(t) = (FEu)(t)$ for every $u \in S_+, t \in \mathbb{R}_+$. For example, if $H : S \to L_\infty(\mathbb{R})$ is given by

$$(Hu)(t) = (Fu)(t) + \limsup_{\tau\to-\infty} u(\tau), \quad t \in \mathbb{R},$$

where F is as in Theorem 5.1, then $(Gu)(t) = (HEu)(t)$ for every $u \in S_+, t \in \mathbb{R}_+$. [26]

5.3.2 A Continuity Property of F_N.

We noted in Section 5.2.2 that G_N has continuity property $\mathcal{C}_{2+}$. In Section 5.4, we need F_N to have continuity property $\mathcal{C}_2$, so that we can show that F_N is L_p-myopic. The following proposition shows that F_N does indeed have continuity property $\mathcal{C}_2$. Again, $L_\infty(\mathbb{R}_+) \subseteq V_+$ and $L_\infty(\mathbb{R}) \subseteq V$, so we can say $G_N : S_+ \to V_+$ and $F_N : S \to V$.

Proposition 5.3: Suppose $G : S_+ \to V_+$ is time-invariant and has approximately finite memory. Let $F : S \to V$ be as in Theorem 5.1. Suppose also that G has continuity property $\mathcal{C}_{p+}$, where $p \geq 1$. Then F has continuity property $\mathcal{C}_p$.

Proof: Choose any $t \in \mathbb{R}$. Let $\varepsilon > 0$. F must have approximately finite memory by (v) of Theorem 5.1, so let Δ be the associate of $\frac{1}{3}\varepsilon$ regarding the approximately finite memory of F. Let δ be the associate of $\frac{1}{3}\varepsilon$ and t regarding the continuity property $\mathcal{C}_{p+}$ of G. Let $s, s' \in S \cap L_p(\mathbb{R})$ such that $\|s - s'\|_p < \delta$. The triangle inequality gives

$$
\begin{aligned}
|(Fs)(t) - (Fs')(t)| \leq |(Fs)(t) - (F\tilde{W}_{t,\Delta}s)(t)| + |(F\tilde{W}_{t,\Delta}s)(t) - (F\tilde{W}_{t,\Delta}s')(t)| \\
+ |(F\tilde{W}_{t,\Delta}s')(t) - (Fs')(t)|. \quad (5.4)
\end{aligned}
$$

Using the approximately finite memory of F, the first and third terms on the right side of (5.4) are bounded by $\frac{1}{3}\varepsilon$. It remains only to bound the second term by $\frac{1}{3}\varepsilon$. Using time-invariance, we have

$$
\begin{aligned}
(F\tilde{W}_{t,\Delta}s)(t) &= (F\tilde{T}_{t-\Delta}\tilde{W}_{t,\Delta}s)(\Delta) \\
&= (F\tilde{W}_{\Delta,\Delta}\tilde{T}_{t-\Delta}s)(\Delta) \\
&= (FEP\tilde{W}_{\Delta,\Delta}\tilde{T}_{t-\Delta}s)(\Delta) \\
&= (GP\tilde{W}_{\Delta,\Delta}\tilde{T}_{t-\Delta}s)(\Delta) \\
&= G_\Delta \sigma,
\end{aligned}
$$

where $\sigma \in S_\Delta$ is given by $\sigma(t) = (\tilde{T}_{t-\Delta}s)(t)$ for $t \in [0, \Delta]$, and where G_Δ is given by (5.1). Similarly,

$$
(F\tilde{W}_{t,\Delta}s')(\Delta) = G_\Delta \sigma',
$$

where $\sigma'(t) = (\tilde{T}_{t-\Delta}s')(t)$ for $t \in [0, \Delta]$. Therefore, the second term on the right side of (5.4) is equal to $|G_\Delta \sigma - G_\Delta \sigma'|$. We have

$$\begin{aligned}
\|\sigma - \sigma'\|_p &= \left(\int_0^\Delta \left| (\tilde{T}_{t-\Delta} s)(\tau) - (\tilde{T}_{t-\Delta} s')(\tau) \right|^p d\tau \right)^{p^{-1}} \\
&= \left(\int_{t-\Delta}^t |s(\tau) - s'(\tau)|^p d\tau \right)^{p^{-1}} \\
&\leq \left(\int_{-\infty}^\infty |s(\tau) - s'(\tau)|^p d\tau \right)^{p^{-1}} = \|s - s'\|_p < \delta.
\end{aligned}$$

So continuity property $\mathcal{C}_{p+}$ gives $|G_\Delta \sigma - G_\Delta \sigma'| < \frac{1}{3}\varepsilon$, and the second term on the right side of (5.4) is also bounded by $\frac{1}{3}\varepsilon$. This completes the proof.

5.3.3 Comments

The results of this section allow us to say there is an $F_N : S \to V$ that is a causal, time-invariant extension (in the sense of Theorem 5.1) of G_N. Further, F_N has approximately finite memory, and has continuity property $\mathcal{C}_2$. We see in Section 5.4 that this is sufficient to show that F_N is L_p-myopic.

5.4 Theorem Showing that F_N is L_p-myopic.

5.4.1 Preliminaries

For $p \geq 1$, let $\mathcal{W}_{L_p}$ be the set of Lebesgue measurable $w : \mathbb{R} \to \mathbb{R}$ such that:

(i) $\int_\mathbb{R} |w(\tau)|^p \, d\tau < \infty$, and

(ii) $\sup_{\tau \in U} |w(\tau)| < \infty$ and $\inf_{\tau \in U} |w(\tau)| > 0$ for each bounded subset U of $\mathbb{R}$.

We say that a time-invariant map $F : S \to V$ is L_p-myopic with respect to $w \in \mathcal{W}_{L_p}$ if for every $\varepsilon > 0$, there is a $\delta > 0$ such that

$$|(Fs)(0) - (Fs')(0)| < \varepsilon \tag{5.5}$$

whenever $s, s' \in S$ and

$$\int_{\mathbb{R}} |w(\tau)[s(\tau) - s'(\tau)]|^p \, d\tau < \delta.$$

This is equivalent[7] to the definition of a myopic map used in [12], which is distinct from the definition used in [44].

We also say that $F : S \to V$ is L-myopic if for every $p \geq 1$ and for every $w \in \mathcal{W}_{L_p}$, F is L_p-myopic with respect to w.

5.4.2 Theorem and Proof

The following theorem shows that F_N is L_p-myopic with respect to every $w \in \mathcal{W}_{L_p}$, for every $p \geq 1$. Therefore, the results in [12] apply to a familiar, reasonable system. More broadly, it gives the relationship between the concepts of a myopic map and of a map with approximately finite memory, and it shows that if a map is L_p-myopic with respect to some $w \in \mathcal{W}_{L_p}$, for some $p \geq 1$, then it is necessarily L_p-myopic with respect to every $w \in \mathcal{W}_{L_p}$, for every $p \geq 1$ (i.e., it is L-myopic).

Theorem 5.2:. Suppose $F : S \to V$ is time-invariant and causal. Then the following statements are equivalent.

(i) F has approximately finite memory, and for some $p \geq 1$, F has continuity property $\mathcal{C}_p$.

(ii) For some $p \geq 1$ and for some $w \in \mathcal{W}_{L_p}$, F is L_p-myopic with respect to w.

[7] The difference between our definition and the definition in [12] (setting $m = 1$ in [12]) is that the range V of F consists of $\mathbb{R}^n$-valued functions, while the range of the analogous maps in [12] consists of $\mathbb{R}$-valued functions. However, because all norms on $\mathbb{R}^n$ are equivalent (specifically, because the Euclidean norm on $\mathbb{R}^n$ in the inequality (5.5) is equivalent to the maximum norm on $\mathbb{R}^n$), F may be treated as a collection of n different maps that are myopic in the sense of [12]. Note that the domain of F is the same as the domain of the analogous maps in [12] — both domains consist of $\mathbb{R}^n$-valued functions. So if a map is L_p-myopic in the sense of this chapter, the approximation results in [12] may be used.

(iii) F has approximately finite memory, and for every $p \geq 1$, F has continuity property $\mathcal{C}_p$.

(iv) F is L-myopic.

This theorem is a direct consequence of Propositions 5.4 and 5.5 below.

Proposition 5.4: Let $p \geq 1$, and choose $w \in \mathcal{W}_{L_p}$. Suppose $F : S \to V$ is time-invariant and causal. Then F is L_p-myopic with respect to w if and only if F has approximately finite memory and continuity property $\mathcal{C}_p$.

Proof: Suppose F has approximately finite memory and continuity property $\mathcal{C}_p$. Choose $\varepsilon > 0$, let Δ be the associate of $\frac{1}{3}\varepsilon$ regarding the approximately finite memory of F, and let δ_0 be the associate at $t = 0$ of $\frac{1}{3}\varepsilon$ regarding the continuity property $\mathcal{C}_p$. Define $c = \inf_{\tau \in [-\Delta, 0]} |w(\tau)|$, and recall from the definition of $\mathcal{W}_{L_p}$ that $c > 0$. Set $\delta = (c\delta_0)^p$. Suppose $s, s' \in S$ with

$$\int_{\mathbb{R}} |w(\tau)[s(\tau) - s'(\tau)]|^p \, d\tau < (c\delta_0)^p = \delta. \tag{5.6}$$

The triangle inequality gives

$$|(Fs)(0) - (Fs')(0)| \leq |(Fs)(0) - (F\tilde{W}_{0,\Delta}s)(0)| + |(F\tilde{W}_{0,\Delta}s)(0) - (F\tilde{W}_{0,\Delta}s')(0)| + |(F\tilde{W}_{0,\Delta}s')(0) - (Fs')(0)|. \tag{5.7}$$

The first and third terms above are bounded by $\frac{1}{3}\varepsilon$ because F has approximately finite memory. It remains to show that the second term is also bounded by $\frac{1}{3}\varepsilon$.

Note that $\tilde{W}_{0,\Delta}S \subseteq S$. Further, $|(\tilde{W}_{0,\Delta}s)(t)|$ and $|(\tilde{W}_{0,\Delta}s')(t)|$ are bounded by b for $\tau \in [t - \Delta, t]$, and are zero elsewhere. So

$$\tilde{W}_{0,\Delta}s, \tilde{W}_{0,\Delta}s' \in S \cap L_p(\mathbb{R}).$$

Thus by continuity property $\mathcal{C}_p$, the second term in (5.7) is bounded by $\frac{1}{3}\varepsilon$ if $\|\tilde{W}_{0,\Delta}s - \tilde{W}_{0,\Delta}s'\|_p < \delta_0$. But since $w(\tau) \geq c$ for $\tau \in [-\Delta, 0]$,

$$\begin{aligned}
\|\tilde{W}_{0,\Delta}s - \tilde{W}_{0,\Delta}s'\|_p &= \left(\int_{-\Delta}^{0} \big|[s(\tau) - s'(\tau)]\big|^p \, d\tau\right)^{p^{-1}} \\
&\leq \frac{w(\tau)}{c}\left(\int_{-\Delta}^{0} \big|[s(\tau) - s'(\tau)]\big|^p \, d\tau\right)^{p^{-1}} \\
&\leq \frac{1}{c}\left(\int_{-\infty}^{\infty} \big|w(\tau)[s(\tau) - s'(\tau)]\big|^p \, d\tau\right)^{p^{-1}} \\
&< \delta_0
\end{aligned}$$

using (5.6). So all three terms on the right side of (5.7) are bounded by $\frac{1}{3}\varepsilon$, and F is L_p-myopic.

Conversely, suppose F is L_p-myopic with respect to w. For each $\beta \in \mathbb{R}$, define $w_\beta : \mathbb{R} \to \mathbb{R}$ by

$$w_\beta(\tau) = \begin{cases} w(\tau), & \tau < -\beta \\ 0, & \tau \geq -\beta \end{cases}.$$

Note $|w_\beta(\tau)| \leq |w(\tau)|$ and $\lim_{\beta\to\infty} |w_\beta(\tau)|^p = 0$ for every $\tau \in \mathbb{R}$. By the Dominated Convergence Theorem (see, for example, Theorem 5.19 of [28]),

$$\lim_{\beta\to\infty} \int_{\mathbb{R}} \big|w_\beta(\tau)\big|^p \, d\tau = 0. \tag{5.8}$$

Now let $\varepsilon > 0$. Let δ is the associate of ε in the L_p-myopic property of F, and let Δ be the associate of δb^{-p} in the limit (5.8). (Recall also that b is the bound on $|s(t)|$ for $s \in S$.) Let $s \in S$, and define $\sigma \in S$ by

$$\sigma(\tau) = \begin{cases} s(\tau), & \tau \leq 0 \\ 0_n, & \tau > 0 \end{cases}.$$

Then

$$\int_{\mathbb{R}} \big|w(\tau)[\sigma(\tau) - (\tilde{W}_{0,\Delta}s)(\tau)]\big|^p \, d\tau = \int_{\mathbb{R}} \big|w_\Delta(\tau)s(\tau)\big|^p \, d\tau$$
$$\leq \int_{\mathbb{R}} \big|w_\Delta(\tau)b\big|^p \, d\tau < \delta.$$

Because F is causal, $(Fs)(0) = (F\sigma)(0)$, which together with the inequality above and the L_p-myopic property of F gives

$$|(Fs)(0) - (F\tilde{W}_{0,\Delta}s)(0)| = |(F\sigma)(0) - (F\tilde{W}_{0,\Delta}s)(0)| < \varepsilon.$$

Thus F has property $\mathcal{A}_0$, and by Proposition 5.2, F has approximately finite memory.

It remains to show that F has continuity property $\mathcal{C}_p$. Fix $t \in \mathbb{R}$, and choose $\varepsilon > 0$. Let δ be the associate of ε in the L_p-myopic property of F, and let Δ be the associate of $\frac{1}{2}\delta(2b)^{-p}$ in the limit (5.8). Set $v_\Delta = w - w_\Delta$. Define $c = \sup_{\tau\in[-\Delta,t]} |w(\tau)|^p$, and note that $c < \infty$ by property (ii) in the definition of $\mathcal{W}_{L_p}$. Also note that $c \geq \sup_{\tau\in[-\Delta,t]} |v_\Delta(\tau)|^p$. Choose any $s, s' \in S \cap L_p(\mathbb{R})$ such that $\|s - s'\|_p < (\frac{1}{2c}\delta)^{\frac{1}{p}}$. Let $\sigma, \sigma' \in S$ be given by

$$\sigma(\tau) = \begin{cases} (\tilde{T}_{-t}s)(\tau), & \tau \leq t \\ 0, & \tau > t \end{cases} \quad \text{and} \quad \sigma'(\tau) = \begin{cases} (\tilde{T}_{-t}s')(\tau), & \tau \leq t \\ 0, & \tau > t \end{cases}.$$

We have

$$\int_{\mathbb{R}} \big|w(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau = \int_{\mathbb{R}} \big|w_\Delta(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau$$
$$+ \int_{\mathbb{R}} \big|v_\Delta(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau. \quad (5.9)$$

Because $|\sigma(t)| \leq b$ and $|\sigma'(t)| \leq b$ for all t, the first term on the right side of (5.9) is bounded by

$$\int_{\mathbb{R}} \big|w_\Delta(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau \le (2b)^p \int_{\mathbb{R}} \big|w_\Delta(\tau)\big|^p \, d\tau < \frac{\delta}{2}$$

using (5.8). The second term on the right side of (5.9) is bounded by

$$\begin{aligned}
\int_{\mathbb{R}} \big|v_\Delta(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau &= \int_{-\Delta}^{t} \big|v_\Delta(\tau)[\sigma(\tau) - \sigma'(\tau)]\big|^p \, d\tau \\
&\le c \int_{-\Delta}^{t} \big|\sigma(\tau) - \sigma'(\tau)\big|^p \, d\tau \\
&= c \int_{(-\Delta-t)}^{0} \big|s(\tau) - s'(\tau)\big|^p \, d\tau \le c(\|s - s'\|_p)^p < \frac{\delta}{2}.
\end{aligned}$$

Now the left side of (5.9) is bounded by δ, and the L_p-myopic property of F gives

$$|(Fs)(0) - (Fs')(0)| = |(F\tilde{T}_{-t}s)(t) - (F\tilde{T}_{-t}s')(t)| = |(F\sigma)(t) - (F\sigma')(t)| < \varepsilon,$$

using the time-invariance and causality of F. So F has continuity property $\mathcal{C}_p$. This completes the proof.

Proposition 5.5: Suppose $F : S \to V$ has approximately finite memory, and suppose for some $p \ge 1$, F has continuity property $\mathcal{C}_p$. Then F also has continuity property $\mathcal{C}_q$, where $q \ge 1$.

Proof: Suppose that F has continuity property $\mathcal{C}_p$, where $p \ge 1$. Let $q \ge 1$. Fix $t \in \mathbb{R}$, and choose $\varepsilon > 0$. Let Δ be the associate of $\frac{1}{3}\varepsilon$ regarding the approximately finite memory of F, and let δ_p be the associate at t of $\frac{1}{3}\varepsilon$ regarding the continuity property $\mathcal{C}_p$. If $p = q$, there is nothing to prove. If $p > q$, define

$$\delta_q = 2b\left(\frac{\delta_p}{2b}\right)^{q^{-1}p},$$

or if $p < q$, define

$$\delta_q = \delta_p \cdot \Delta^{(q^{-1}-p^{-1})}.$$

Suppose that $s, s' \in S \cap L_q(\mathbb{R})$. The triangle inequality gives

$$\begin{aligned} |(Fs)(t) - (Fs')(t)| \le |(Fs)(t) - (F\tilde{W}_{t,\Delta}s)(t)| + |(F\tilde{W}_{t,\Delta}s)(t) - (F\tilde{W}_{t,\Delta}s')(t)| \\ + |(F\tilde{W}_{t,\Delta}s')(t) - (Fs')(t)|. \quad (5.10) \end{aligned}$$

The first and third terms above are bounded by $\frac{1}{3}\varepsilon$ because F has approximately finite memory. It remains to show that the second term is also bounded by $\frac{1}{3}\varepsilon$. Because F has continuity property $\mathcal{C}_p$, we seek to bound $\|\tilde{W}_{t,\Delta}s - \tilde{W}_{t,\Delta}s'\|_p$ by δ_p.

Suppose $p > q$. Then

$$\begin{aligned} \|\tilde{W}_{t,\Delta}s - \tilde{W}_{t,\Delta}s'\|_p &= \left(\int_{t-\Delta}^{t} \left|s(\tau) - s'(\tau)\right|^p d\tau\right)^{p^{-1}} \\ &= 2b\left(\int_{t-\Delta}^{t} \left|\frac{s(\tau) - s'(\tau)}{2b}\right|^p d\tau\right)^{p^{-1}}. \end{aligned} \quad (5.11)$$

Because $|s(\tau) - s'(\tau)| \le 2b$ and $p > q$, we have

$$\left|\frac{s(\tau) - s'(\tau)}{2b}\right|^p \le \left|\frac{s(\tau) - s'(\tau)}{2b}\right|^q,$$

and by (5.11),

$$\begin{aligned} \|\tilde{W}_{t,\Delta}s - \tilde{W}_{t,\Delta}s'\|_p &\le 2b\left(\int_{t-\Delta}^{t} \left|\frac{s(\tau) - s'(\tau)}{2b}\right|^q d\tau\right)^{p^{-1}} \\ &= \frac{2b}{(2b)^{p^{-1}q}}\left[\left(\int_{t-\Delta}^{t} \left|s(\tau) - s'(\tau)\right|^q d\tau\right)^{q^{-1}}\right]^{p^{-1}q} \\ &\le \frac{2b}{(2b)^{p^{-1}q}}\left[\left(\int_{-\infty}^{\infty} \left|s(\tau) - s'(\tau)\right|^q d\tau\right)^{q^{-1}}\right]^{p^{-1}q} \\ &= \frac{2b}{(2b)^{p^{-1}q}}(\|s - s'\|_q)^{p^{-1}q} < \frac{2b}{(2b)^{p^{-1}q}}\ (\delta_q)^{p^{-1}q} = \delta_p\,. \end{aligned}$$

Because F has continuity property $\mathcal{C}_p$, the second term in (5.10) is bounded by $\frac{1}{3}\varepsilon$, and F has continuity property $\mathcal{C}_q$.

Suppose now that $p < q$. It is known ([28], exercise 5, p. 143) that if $f : [t-\Delta, t] \to \mathbb{R}$ with $|f|^p$ and $|f|^q$ Lebesgue integrable,

$$\left(\frac{1}{\Delta}\int_{t-\Delta}^{t} |f(\tau)|^p \, d\tau\right)^{p^{-1}} \leq \left(\frac{1}{\Delta}\int_{t-\Delta}^{t} |f(\tau)|^q \, d\tau\right)^{q^{-1}} .$$

Therefore, with $f(\tau) = |(\tilde{W}_{t,\Delta}s)(\tau) - (\tilde{W}_{t,\Delta}s')(\tau)|$ for $\tau \in [t-\Delta, t]$,

$$\begin{aligned}\left(\frac{1}{\Delta}\int_{t-\Delta}^{t} |(\tilde{W}_{t,\Delta}s)(\tau) - (\tilde{W}_{t,\Delta}s')(\tau)|^p \, d\tau\right)^{p^{-1}} \\ \leq \left(\frac{1}{\Delta}\int_{t-\Delta}^{t} |(\tilde{W}_{t,\Delta}s)(\tau) - (\tilde{W}_{t,\Delta}s')(\tau)|^q \, d\tau\right)^{q^{-1}} . \quad (5.12)\end{aligned}$$

The left side of (5.12) is equal to

$$\left(\frac{1}{\Delta}\int_{-\infty}^{\infty} |(\tilde{W}_{t,\Delta}s)(\tau) - (\tilde{W}_{t,\Delta}s')(\tau)|^p \, d\tau\right)^{p^{-1}}$$

and the right side of (5.12) is no larger than

$$\left(\frac{1}{\Delta}\int_{-\infty}^{\infty} |s(\tau) - s'(\tau)|^q\right)^{q^{-1}}$$

which gives

$$\|\tilde{W}_{t,\Delta}s - \tilde{W}_{t,\Delta}s'\|_p \leq \Delta^{(p^{-1}-q^{-1})}\|s-s'\|_q < \Delta^{(p^{-1}-q^{-1})}\delta_q < \delta_p \, .$$

Because F has continuity property $\mathcal{C}_p$, the second term in (5.10) is bounded by $\frac{1}{3}\varepsilon$, and F again has continuity property $\mathcal{C}_q$. This completes the proof.

We may now prove Theorem 5.2.

Proof of Theorem 6: By Proposition 5.5, (i) implies (iii). By Proposition 5.4, (iii) implies (iv). Clearly (iv) implies (ii), and by Proposition 5.4 again, (ii) implies (i). So (i), (ii), (iii), and (iv) are equivalent, and the proof is complete.

5.4.3 Conclusions

In Section 5.2.2, we showed that if G_N is drawn from a familiar class of feedback systems, then it has approximately finite memory, it is time-invariant, and it has continuity property $\mathcal{C}_{2+}$. Proposition 5.3 and Theorem 5.2 show that F_N is myopic in the sense of [12], where F_N is an extension of G_N, in the sense of Theorem 5.1. This establishes the applicability of the approximation results in [12].

Appendix A

Additional Proofs Relating to Material in Chapter 4

A.1 Proof of Proposition 4.1

We will need the following definitions for the proof of Proposition 4.1. For any open interval $(a, b) \subseteq \mathbb{R}$ and any map $f : (a, b) \to \mathbb{R}$, we say q is a "point of approximate continuity" of f if for each $\varepsilon > 0$,

$$\lim_{r \to 0} \frac{\mu\big(\bar{B}_r(q) \bigcap \{\xi \in (a, b) : |f(\xi) - f(q)| \geq \varepsilon\}\big)}{\mu(\bar{B}_r(q))} = 0,$$

where μ denotes Lebesgue measure. We will say a finite set of points $\{q_0, \ldots, q_k\} \subset (a, b)$ is an "approximately continuous partition" for f of (a, b) if each q_k is a point of approximate continuity of f, and if $q_j < q_{j+1}$ for $j = 0, \ldots, k - 1$. Let $\mathcal{Q}(f; a, b)$ be the set of all approximately continuous partitions for f of (a, b). The function f is said to have essentially bounded variation on (a, b) if

$$\sup_{Q \in \mathcal{Q}(f;a,b)} \sum_{j=1}^{k_Q} \big|f(q_j) - f(q_{j-1})\big|$$

is finite (where $\{q_0, \ldots, q_{k_Q}\} = Q$). In this case, we say that

$$V_E(f) = \sup_{Q \in \mathcal{Q}(f;a,b)} \sum_{j=1}^{k_Q} \big| f(q_j) - f(q_{j-1}) \big|$$

is the "essential variation" of f on (a, b).

For any positive integer ψ and any open $\mathcal{O} \subseteq \mathbb{R}^\psi$, let $C_c^1(\mathcal{O}, \mathbb{R}^\psi)$ be the set of functions from $\mathcal{O}$ to $\mathbb{R}^\psi$ which have compact support, and for which every first-order partial derivative exists and is continuous. If $\phi \in C_c^1(\mathcal{O}, \mathbb{R}^\psi)$, we write the divergence of ϕ as $\operatorname{div} \phi$. A function $v \in L_p^1(\mathcal{O})$ is said to have bounded variation on $\mathcal{O}$ in the sense of distributions if

$$\sup \bigg\{ \int_{\mathcal{O}} v(x) \operatorname{div} \phi(x)\, dx : \phi \in C_c^1(\mathcal{O}, \mathbb{R}^\psi),\ \sup_{x \in \mathcal{O}} |\phi(x)|_2 \leq 1 \bigg\}$$

is finite, in which case we say

$$V_S(v) = \sup \bigg\{ \int_{\mathcal{O}} v(x) \operatorname{div} \phi(x)\, dx : \phi \in C_c^1(\mathcal{O}, \mathbb{R}^\psi),\ \sup_{x \in \mathcal{O}} |\phi(x)|_2 \leq 1 \bigg\} \qquad \text{(A.1)}$$

is the "total distributional variation" on $\mathcal{O}$ of v. We also define the norm $\| \cdot \|_{BV}$ on such functions by

$$\|v\|_{BV} = \|v\|_1 + V_S(v).$$

For any Lebesgue measurable $\mathcal{O} \subseteq \mathbb{R}^m$ such that $\mathcal{O} \supseteq [0,1]^m$, let $E_{\mathcal{O}}$ be the map from $L_p^1([0,1]^m)$ to $L_p^1(\mathcal{O})$ given for each $v \in L_p^1([0,1]^m)$ by

$$(E_{\mathcal{O}} v)(x) = \begin{cases} v(x), & x \in [0,1]^m \\ 0, & \text{otherwise} \end{cases}, x \in \mathcal{O}.$$

Proof of Proposition 4.1: First we will use the proof of Theorem 2 of Section

5.10.2 of [55]. Let $D = (-0.1, 1.1)$, and let $\mathcal{O} = D^m$. Note that $\mathcal{O}$ is a bounded open set containing $[0,1]^m$. For the moment, fix $u \in U$, $k \in \{1,\dots,\ell\}$, $i \in \{1,\dots,m\}$, and $y \in D^{m-1}$. If $y \notin [0,1]^{m-1}$, then $(E_{\mathcal{O}}u_k)(\rho_m(\tau,i,y)) = 0$ for every $\tau \in D$, and therefore

$$V_E\big((E_{\mathcal{O}}u_k)(\rho_m(\cdot,i,y))\big) = 0 \qquad \text{if } y \notin [0,1]^{m-1}. \tag{A.2}$$

If $y \in [0,1]^{m-1}$, we claim that the essential variation of $(E_{\mathcal{O}}u_k)(\rho_m(\cdot,i,y))$ on D is bounded, and specifically that

$$V_E\big((E_{\mathcal{O}}u_k)(\rho_m(\cdot,i,y))\big) \leq 2b + \bar{\Phi} \qquad \text{if } y \in [0,1]^{m-1}. \tag{A.3}$$

In order to establish the claim, choose any approximately continuous partition $Q = \{q_0,\dots,q_{k_Q}\}$ for $(E_{\mathcal{O}}u_k)(\rho_m(\cdot,i,y))$ of D. Let α be the smallest integer such that $q_\alpha \geq 0$, and let β be the largest integer such that $q_\beta \leq 1$. Construct a partition $P = \{p_0,\dots,p_{k_P}\}$ of $[0,1]$ consisting of 0, 1, and all elements of Q between 0 and 1. Now $\{q_\alpha,\dots,q_\beta\} \subseteq P$ (with equality if $q_\alpha = 0$ and $q_\beta = 1$), and $\rho_m(q_j,i,y) \in [0,1]^m$ for $\alpha \leq j \leq \beta$, so

$$\begin{aligned}\sum_{j=\alpha+1}^{\beta} &\big|(E_{\mathcal{O}}u_k)(\rho_m(q_j,i,y)) - (E_{\mathcal{O}}u_k)(\rho_m(q_{j-1},i,y))\big| \\ &\leq \sum_{j=1}^{k_P}\big|u_k(\rho_m(p_j,i,y)) - u_k(\rho_m(p_{j-1},i,y))\big| \leq \Phi(u_k,i,y) \leq \bar{\Phi}.\end{aligned} \tag{A.4}$$

Further, because $(E_{\mathcal{O}}u_k)(\rho_m(\cdot,i,y))$ is zero except on $[0,1]$, and because $\|u\|_\infty \leq b$, we have

$$\sum_{j=1}^{\alpha}\big|(E_{\mathcal{O}}u_k)(\rho_m(q_j,i,y)) - (E_{\mathcal{O}}u_k)(\rho_m(q_{j-1},i,y))\big| = \big|(E_{\mathcal{O}}u_k)(\rho_m(q_\alpha,i,y))\big| \leq b, \tag{A.5}$$

and similarly

$$\sum_{j=\beta+1}^{k_Q} \big|(E_{\mathcal{O}} u_k)(\rho_m(q_j,i,y)) - (E_{\mathcal{O}} u_k)(\rho_m(q_{j-1},i,y))\big| = \big|(E_{\mathcal{O}} u_k)(\rho_m(q_\beta,i,y))\big| \le b. \tag{A.6}$$

Adding (A.4), (A.5), and (A.6) gives

$$\sum_{j=1}^{k_Q} \big|(E_{\mathcal{O}} u_k)(\rho_m(q_j,i,y)) - (E_{\mathcal{O}} u_k)(\rho_m(q_{j-1},i,y))\big| \le 2b + \bar{\Phi}.$$

This holds for every Q, so we have established the claim (A.3).

Now suppose $\phi \in C_c^1(\mathcal{O}, \mathbb{R}^m)$. Then

$$\begin{aligned} &\int_{\mathcal{O}} (E_{\mathcal{O}} u_k)(x) \operatorname{div} \phi(x) dx \\ &\qquad = \sum_{i=1}^{m} \int_{D^{m-1}} \int_D \big[(E_{\mathcal{O}} u_k)(\rho_m(x_i,i,y))\big] \left[\frac{\partial \phi_i}{\partial x_i}(\rho_m(x_i,i,y))\right] dx_i dy. \end{aligned} \tag{A.7}$$

But Theorem 1 of Section 5.10.1 of [55] guarantees that

$$V_S\big((E_{\mathcal{O}} u_k)(\rho_m(\cdot,i,y))\big) = V_E\big((E_{\mathcal{O}} u_k)(\rho_m(\cdot,i,y))\big).$$

(Here, $\psi = 1$ in (A.1)). Since for each i we have $\phi_i(\rho_m(\cdot,i,y)) \in C_c^1(D,\mathbb{R})$ and $|\phi_i(\rho_m(t,i,y))| \le 1$ for every $t \in D$, it follows that

$$\int_D \big[(E_{\mathcal{O}} u_k)(\rho_m(x_i,i,y))\big] \left[\frac{\partial \phi_i}{\partial x_i}(\rho_m(x_i,i,y))\right] dx_i \le V_E\big((E_{\mathcal{O}} u_k)(\rho_m(\cdot,i,y))\big)$$

for each i and each $y \in D^{m-1}$. Therefore (A.7) implies that

$$\int_{\mathcal{O}} (E_{\mathcal{O}} u_k)(x) \operatorname{div} \phi(x) dx \le \sum_{i=1}^{m} \int_{D^{m-1}} V_E\big((E_{\mathcal{O}} u_k)(\rho_m(\cdot,i,y))\big) dy.$$

Now using (A.2) and (A.3), we have

$$\int_{\mathcal{O}} (E_{\mathcal{O}} u_k)(x) \operatorname{div} \phi(x) dx \leq \sum_{i=1}^{m} \int_{[0,1]^{m-1}} V_E\big((E_{\mathcal{O}} u_k)(\rho_m(\cdot, i, y))\big) dy$$
$$\leq 2bm + m\bar{\Phi}.$$

Because this holds for any $\phi \in C_c^1(\mathcal{O}, \mathbb{R}^m)$,

$$V_S(E_{\mathcal{O}} u_k) \leq 2bm + m\bar{\Phi}.$$

(Here, $\psi = m$ in (A.1)). Further, $\|E_{\mathcal{O}} u_k\|_1 \leq b$, so

$$\|E_{\mathcal{O}} u_k\|_{BV} \leq b + 2mb + m\bar{\Phi}. \tag{A.8}$$

This holds for every every $u \in U$ and every $k = 1, \ldots, \ell$.

For $k = 1, \ldots, \ell$, let Υ_k and U_k denote $\{E_{\mathcal{O}} u_k : u \in U\}$ and $\{u_k : u \in U\}$, respectively. Fix k, and let $\{v_j\}$ be a sequence in U_k.[1] Now $\{E_{\mathcal{O}} v_j\}$ is a sequence in Υ_k such that for every j, by (A.8),

$$\|E_{\mathcal{O}} v_j\|_{BV} \leq b + 2mb + m\bar{\Phi}.$$

Further, $\mathcal{O}$ has a "Lipschitz boundary" in the sense of Section 4.2 of [55]. Therefore by Theorem 4 of Section 5.2.3 of [55], there is a subsequence $\{E_{\mathcal{O}} v_j\}_{j \in J}$ of $\{E_{\mathcal{O}} v_j\}$ which converges in $L_1^1(\mathcal{O})$. The corresponding subsequence $\{v_j\}_{j \in J}$ of $\{v_j\}$ must converge in $L_1^1([0,1]^m)$. But $\|u\|_\infty \leq b$ for each $u \in U$, and consequently $\|v_j\|_\infty \leq b$ for each j. It follows that $\|v_j\|_p \leq (b^{p-1}\|v_j\|_1)^{\frac{1}{p}}$ for each j, so the subsequence $\{v_j\}_{j \in J}$ converges also in $L_p^1([0,1]^m)$. Therefore each U_k is relatively compact in

[1] We emphasize that each v_j is an element of U_k, and therefore of $L_p^1([0,1]^m)$. To clarify, for any $x \in [0,1]^m$, $v_j(x)$ is a real number, not a vector. The "j" in v_j denotes its order in a sequence, while the "k" in u_k denotes a vector component of an element of $L_p^\ell([0,1]^m)$.

$L^1_p([0,1]^m)$, and consequently each $\mathrm{cl}(U_k)$ (i.e. the closure of each U_k) is compact.

By Tychonoff's Theorem (Theorem 5 of Section I.8 of [48], for example), $\prod_{k=1}^{\ell} \mathrm{cl}(U_k)$ is compact in $L^\ell_p([0,1]^m)$ under the product topology. A metric for the product topology (also given in Section I.8 of [48], for example) is

$$d(u,u') = \sum_{k=1}^{\ell} \frac{\|u_k - u'_k\|_p}{2^k\big(1 + \|u_k - u'_k\|_p\big)}, \qquad u, u' \in L^\ell_p([0,1]^m),$$

where $\|\cdot\|_p$ denotes the norm on $L^1_p([0,1]^m)$. It is easily seen that whenever $d(u,u') < \frac{\ell}{2^{\ell+1}}$,

$$d(u,u') \geq \frac{1}{2^{\ell+1}} \sum_{k=1}^{\ell} \|u_k - u'_k\|_p. \tag{A.9}$$

For each $u \in L^\ell_p([0,1]^m)$ and each $k = 1,\ldots,\ell$, let $\hat{u}_k$ represent the element of $L^\ell_p([0,1]^m)$ such that for each $x \in [0,1]^m$, the components of $\hat{u}_k(x) \in \mathbb{R}^\ell$ are given by

$$\big(\hat{u}_k(x)\big)_\psi = \begin{cases} u_k(x), & k = \psi, \\ 0, & \text{otherwise.} \end{cases}, \qquad \psi = 1,\ldots,\ell.$$

Using this notation, clearly

$$\|u_k - u'_k\|_p = \|\hat{u}_k - \hat{u}'_k\|_p,$$

where the norms on the left and right are on $L^1_p([0,1]^m)$ and $L^\ell_p([0,1]^m)$, respectively. Further,

$$u - u' = \sum_{k=1}^{\ell} \hat{u}_k - \hat{u}'_k.$$

So using the triangle inequality on (A.9) tells us that

$$d(u,u') \geq \frac{1}{2^{\ell+1}} \|u - u'\|_p$$

whenever $d(u,u') < \frac{\ell}{2^{\ell+1}}$. Therefore a sequence that converges under the product topology must also converge under the norm topology on $L_p^\ell(\mathbb{R}^m)$. It follows that $\prod_{k=1}^{\ell} \mathrm{cl}(U_k)$ is relatively compact in the norm topology. Now U is a subset of $\prod_{k=1}^{\ell} \mathrm{cl}(U_k)$, so U is relatively compact in $L_p^\ell([0,1]^m)$, and the proof is complete.

A.2 Proof of Proposition 4.2

To make the following proof more readable, we use the following notation for open balls in $\mathbb{R}^2$. For every $r > 0$ and every $x \in \mathbb{R}^2$ define $B(x,r) = B_r(x)$.

Proof of Proposition 4.2: We prove this theorem by contradiction. Suppose that f, r, and $\Theta(f,r)$ are as in the statement of Proposition 4.2, but $\mu(\Theta(f,r)) > \pi r^2 + 2r\bar{\Lambda}(f)$. Let

$$\varepsilon = \mu(\Theta(f,r)) - (\pi r^2 + 2r\bar{\Lambda}(f)), \tag{A.10}$$

and note that $\varepsilon > 0$.

To set up the contradiction, let $\zeta : (0,1) \to \mathbb{R}$ be defined by

$$\zeta(q) = \frac{\sin^{-1}(q)}{q} - 1, \quad q \in (0,1).$$

Using L'Hospital's rule,[2] we can see that

$$\lim_{q\to 0} \frac{\sin^{-1}(q)}{q} = 1,$$

and therefore

$$\lim_{q\to 0} \zeta(q) = 0. \tag{A.11}$$

By (A.11), noticing that ζ is positive-valued on $(0,1)$, we can see that there is a

[2]For the reader's convenience, we note that the derivative of the inverse sine function is given by $\frac{d}{dq}\sin^{-1} q = (\sqrt{1-q^2})^{-\frac{1}{2}}$ for $q \in (-1,1)$.

$q_0 > 0$ such that

$$\zeta(q) < \frac{\varepsilon}{8\bar{\Lambda}(f)r} \quad \text{for every } 0 < q < q_0. \tag{A.12}$$

Since $\varepsilon > 0$, we may choose $\delta > 0$ such that the following five conditions hold:

(i) $2\pi r\delta < \frac{1}{4}\varepsilon$,

(ii) $\pi\delta^2 < \frac{1}{4}\varepsilon$,

(iii) $2\delta\bar{\Lambda}(f) < \frac{1}{4}\varepsilon$,

(iv) $\delta < r$, and

(v) $\delta < rq_0$.

Now let k be an integer such that

$$k > \frac{1}{2\delta}\bar{\Lambda}(f). \tag{A.13}$$

We will twice make use of the following fact: Suppose $\tau, \tau' \in [0, 1]$ with $\tau > \tau'$. Because $\{\tau', \tau\}$ is a partition for $[\tau', \tau]$ (albeit a rather trivial partition), $|f(\tau) - f(\tau')|_2$ can be no larger than the arc length of the restriction of f to $[\tau', \tau]$. But this arc length is $\bar{\Lambda}(f, \tau) - \bar{\Lambda}(f, \tau')$ by Theorem 6.18 of [46]. Therefore

$$|f(\tau) - f(\tau')|_2 \leq \bar{\Lambda}(f, \tau) - \bar{\Lambda}(f, \tau'), \qquad 0 \leq \tau' < \tau \leq 1. \tag{A.14}$$

Let $t_0 = 0$ and $t_k = 1$, and note that $\bar{\Lambda}(f, t_0) = 0$ and $\bar{\Lambda}(f, t_k) = \bar{\Lambda}(f)$. By Theorem 6.19 of [46], $\bar{\Lambda}(f, \cdot)$ is continuous and increasing, so we can find $t_1, \ldots, t_{k-1}$ such that $\bar{\Lambda}(f, t_i) = \frac{i}{k}\bar{\Lambda}(f)$ for $i = 1, \ldots, k-1$. Define

$$d_i = |f(t_i) - f(t_{i-1})|_2, \qquad i = 1, \ldots, k,$$

so that from the definition of arc length,

$$\sum_{i=1}^{k} d_i \leq \bar{\Lambda}(f). \tag{A.15}$$

Also note that using (A.13) and (A.14),

$$d_i \leq \bar{\Lambda}(f, t_i) - \bar{\Lambda}(f, t_{i-1}) = \frac{\bar{\Lambda}(f)}{k} < 2\delta \tag{A.16}$$

for $i = 1, \dots k$.

Suppose $x \in \Theta(f, r)$. Then there is a $\tau_x \in [0, 1]$ such that

$$|f(\tau_x) - x|_2 < r. \tag{A.17}$$

By "rounding," we may choose an integer j such that

$$\left| j - \frac{k\bar{\Lambda}(f, \tau_x)}{\bar{\Lambda}(f)} \right| \leq \frac{1}{2}.$$

Then using (A.16),

$$|\bar{\Lambda}(f, t_j) - \bar{\Lambda}(f, \tau_x)| = \left| \frac{j\bar{\Lambda}(f)}{k} - \bar{\Lambda}(f, \tau_x) \right| \leq \frac{\bar{\Lambda}(f)}{2k} < \delta.$$

It follows from (A.14) that

$$|f(t_j) - f(\tau_x)|_2 < \delta.$$

Now using the triangle inequality on the above and (A.17), we have that

$$|f(t_j) - x|_2 < \rho,$$

where $\rho = r + \delta$. Therefore

$$\Theta(f, r) \subseteq \bigcup_{i=0}^{k} B(f(t_i), \rho) \tag{A.18}$$

We now consider the Lebesgue measure of $\Theta(f, r)$. We have from (A.18) that

$$\mu(\Theta(f, r)) \leq \mu\Big(\bigcup_{i=0}^{k} B(f(t_i), \rho)\Big).$$

But

$$\bigcup_{i=0}^{k} B(f(t_i), \rho) = \bigcup_{i=0}^{k}\Big(B(f(t_i), \rho) - \bigcup_{j=0}^{i-1} B(f(t_j), \rho)\Big),$$

where the right side of the above is a disjoint union. Therefore,

$$\begin{aligned} \mu(\Theta(f, r)) &\leq \sum_{i=0}^{k} \mu\Big(B(f(t_i), \rho) - \bigcup_{j=0}^{i-1} B(f(t_j), \rho)\Big) \\ &\leq \mu\big(B(f(t_0), \rho)\big) + \sum_{i=1}^{k} \mu\Big(B(f(t_i), \rho) - \bigcup_{j=0}^{i-1} B(f(t_j), \rho)\Big) \\ &\leq \mu\big(B(f(t_0), \rho)\big) + \sum_{i=1}^{k} \mu\Big(B(f(t_i), \rho) - B(f(t_{i-1}), \rho)\Big). \end{aligned}$$

Since $\rho > \delta$, it is clear from (A.16) that $B(f(t_i), \rho)$ and $B(f(t_{i-1}), \rho)$ intersect, for $i = 1, \ldots, k$. The part of $B(f(t_i), \rho)$ that does not belong to $B(f(t_{i-1}), \rho)$ has measure

$$\mu\Big(B(f(t_i), \rho) - B(f(t_{i-1}), \rho)\Big) = d_i \rho \sqrt{1 - \big(\tfrac{d_i}{2\rho}\big)^2} + 2\rho^2 \sin^{-1}\big(\tfrac{d_i}{2\rho}\big).$$

Therefore since $\mu(B(f(t_0), \rho)) = \pi\rho^2$,

$$\begin{aligned}
\mu(\Theta(f,r)) &\leq \pi\rho^2 + \sum_{i=1}^{k} \left(d_i\rho\sqrt{1 - \left(\tfrac{d_i}{2\rho}\right)^2} + 2\rho^2 \sin^{-1}\left(\tfrac{d_i}{2\rho}\right) \right) \\
&\leq \pi\rho^2 + \sum_{i=1}^{k} d_i\rho \left(\sqrt{1 - \left(\tfrac{d_i}{2\rho}\right)^2} + \tfrac{2\rho}{d_i} \sin^{-1}\left(\tfrac{d_i}{2\rho}\right) \right) \\
&\leq \pi\rho^2 + \sum_{i=1}^{k} d_i\rho \left(1 + \tfrac{2\rho}{d_i} \sin^{-1}\left(\tfrac{d_i}{2\rho}\right) \right),
\end{aligned}$$

recalling that $d_i < 2\delta$ and $\rho = r + \delta$, so that $d_i < 2\rho$.

We can now write

$$\begin{aligned}
\mu(\Theta(f,r)) &\leq \pi\rho^2 + \sum_{i=1}^{k} d_i\rho \left(2 + \zeta(\tfrac{d_i}{2\rho}) \right) \\
&\leq \pi r^2 + 2\pi r\delta + \pi\delta^2 + 2r\bar{\Lambda}(f) + 2\delta\bar{\Lambda}(f) + \sum_{i=1}^{k} d_i\rho\zeta(\tfrac{d_i}{2\rho}),
\end{aligned}$$

where we have used (A.15) and $\rho = r + \delta$. Substituting (A.10), we have

$$\varepsilon \leq 2\pi r\delta + \pi\delta^2 + 2\delta\bar{\Lambda}(f) + \sum_{i=1}^{k} d_i\rho\zeta(\tfrac{d_i}{2\rho}). \tag{A.19}$$

We return to the five conditions we placed on δ. By condition (v), $\delta < rq_0 < \rho q_0$, and using (A.16), $\frac{d_i}{2\rho} < q_0$. Then condition (iv) and (A.12) give us

$$\zeta(\tfrac{d_i}{2\rho}) < \frac{\varepsilon}{8\bar{\Lambda}(f)r} < \frac{\varepsilon}{4(r+\delta)\bar{\Lambda}(f)} = \frac{\varepsilon}{4\rho\bar{\Lambda}(f)}, \quad i = 1, \ldots, k.$$

Now using (A.15), we can write

$$\sum_{i=1}^{k} d_i\rho\zeta(\tfrac{d_i}{2\rho}) < \frac{\varepsilon}{4\bar{\Lambda}(f)} \sum_{i=1}^{k} d_i \leq \frac{\varepsilon}{4}. \tag{A.20}$$

Together, (A.20) and conditions (i), (ii), and (iii) contradict (A.19). Therefore

$\mu(\Theta(f,r)) \leq \pi r^2 + 2r\bar{\Lambda}(f)$, and the proof is complete.

A.3 Proof of Proposition 4.3

We will first prove some auxiliary propositions, which will be used in the proof of Proposition 4.3.

Proposition A.1: Let $f : [0,1] \to \mathbb{R}^2$ be given by

$$f(t) = ty + (1-t)x, \quad t \in [0,1]$$

(i.e. f is the rectifiable curve specifying a straight line from x to y). If $\theta_1, \theta_2, \theta_3 \in [0,1]$ with $\theta_1 \leq \theta_2 \leq \theta_3$, then

$$|f(\theta_1) - f(\theta_3)|_2 = |f(\theta_1) - f(\theta_2)|_2 + |f(\theta_2) - f(\theta_3)|_2. \tag{A.21}$$

Proof: Consider that

$$\begin{aligned}
|f(\theta_1) - f(\theta_2)|_2 &= \left|\theta_1 y + (1-\theta_1)x - \theta_2 y - (1-\theta_2)x\right|_2 \\
&= \left|(\theta_1 - \theta_2)y - (\theta_1 - \theta_2)x\right|_2 \\
&= |\theta_1 - \theta_2|\,|x-y|_2 \\
&= (\theta_2 - \theta_1)\,|x-y|_2
\end{aligned}$$

since $\theta_2 > \theta_1$. By the same reasoning, $|f(\theta_2) - f(\theta_3)|_2 = (\theta_3 - \theta_2)\,|x-y|_2$ and $|f(\theta_1) - f(\theta_3)|_2 = (\theta_3 - \theta_1)\,|x-y|_2$. Then (A.21) follows from addition. This completes the proof.

We use the following notation for open balls in $[0,1]^2$. For every $r > 0$ and every $x \in \mathbb{R}^2$, define $B'(x,r) = [0,1]^2 \bigcap B_r(x)$.

Proposition A.2: Suppose A is a closed subset of $[0,1]^2$, and set $A' = [0,1]^2 - A$. Assume $x, y \in A'$, and let $u : [0,1]^2 \to \mathbb{R}^\ell$ be locally Lipschitz on A', with Lipschitz constant λ. Suppose $f : [0,1] \to \mathbb{R}^2$ is given by

$$f(t) = ty + (1-t)x, \quad t \in [0,1]$$

(i.e. f is the rectifiable curve specifying a straight line from x to y), and suppose further that the graph of f is contained entirely within A', i.e. $f([0,1]) \subseteq A'$. Then

$$|u(x) - u(y)|_1 \leq \lambda |x - y|_2.$$

Proof: Note that $A' \subseteq [0,1]^2$ is open, as a subset of $[0,1]^2$ (though not necessarily as a subset of $\mathbb{R}^2$). So for each $t \in [0,1]$, we may pick $r(t) > 0$ such that the restriction of u to $B'(f(t), r(t))$ is Lipschitz, with Lipschitz constant λ. The set

$$\{B'(f(t), r(t)) : t \in [0,1]\} \tag{A.22}$$

is an open cover for $f([0,1])$, the set of points on the straight line from x to y. Because $f([0,1])$ is the continuous image of a compact set, it is compact (see [46], Theorem 4.25 or Section 6.9). There is therefore a finite sub-cover for $f([0,1])$ composed of sets from (A.22). In fact, using Zorn's lemma (Theorem 4.1-6 of [56], for example) with a partial ordering defined by inclusion, there is a minimal such covering, i.e., there is a positive integer k and a finite set $\{t_1, \ldots, t_k\}$ (assume $t_1 < t_2 < \cdots < t_k$) such that

$$\{B'(f(t_i), r(t_i)) : i = 1, \ldots, k\} \tag{A.23}$$

covers $f([0,1])$, and such that for every $j = 1, \ldots, k$,

$$\{B'(f(t_i), r(t_i)) : i = 1, \ldots, k; i \neq j\} \tag{A.24}$$

does not cover $f([0,1])$. For $i = 1, \ldots, k$, let

$$\rho_i = \frac{r(t_i)}{|x - y|_2}$$
$$I_i = (t_i - \rho_i, t_i + \rho_i) \bigcap [0,1].$$

Note that

$$f(I_i) = B'(f(t_i), r(t_i)) \bigcap f([0,1]). \tag{A.25}$$

Because f is a one-to-one mapping, it follows from (A.23) and (A.24) that

$$\{I_i : i = 1, \ldots, k\} \tag{A.26}$$

covers $[0,1]$, and that for $j = 1, \ldots, k$,

$$\{I_i : i = 1, \ldots, k; i \neq j\} \tag{A.27}$$

does not cover $[0,1]$.

Now set $\sigma_0 = 0$ and $\sigma_k = 1$, and for $i = 1, \ldots, k-1$, let

$$\sigma_i = \tfrac{1}{2}(t_i + \rho_i) + \tfrac{1}{2}(t_{i+1} + \rho_{i+1}).$$

It is not difficult to see that

$$\sigma_i, \sigma_{i+1} \in I_{i+1}, \quad i = 0, \ldots, k-1, \tag{A.28}$$

and also that

$$\sigma_0 < \sigma_1 < \cdots < \sigma_k. \tag{A.29}$$

By the triangle inequality,

$$|u(x) - u(y)|_1 \leq \sum_{i=0}^{k-1} \big|u(f(\sigma_i)) - u(f(\sigma_{i+1}))\big|_1.$$

Using (A.25) and (A.28), $f(\sigma_i)$ and $f(\sigma_{i+1})$ are both in $B'(f(t_{i+1}), r(t_{i+1}))$ for $i = 0, \ldots, k-1$, so

$$|u(x) - u(y)|_1 \leq \sum_{i=0}^{k-1} \lambda \big|f(\sigma_i) - f(\sigma_{i+1})\big|_2.$$

Finally, by (A.29), we may use Proposition A.1 repeatedly to obtain

$$|u(x) - u(y)|_1 \leq \lambda \big|f(\sigma_0) - f(\sigma_k)\big|_2 = \lambda \big|x - y\big|_2.$$

This completes the proof.

We are now prepared to prove Proposition 4.3.

Proof of Proposition 4.3: Assume u is locally Lipschitz on $A' = [0,1]^2 - A$ with Lipschitz constant λ. Let $x, x' \in F$. We must show that

$$|u(x) - u(x')|_1 \leq \lambda |x - x'|_2. \tag{A.30}$$

Let $f : [0,1] \to \mathbb{R}^2$ be given by

$$f(t) = tx' + (1-t)x, \quad t \in [0,1]$$

(i.e. f is the rectifiable curve specifying a straight line from x to x'). By Proposition A.2, if $f(t) \in A'$ for every $t \in [0,1]$, then (A.30) holds.

On the other hand, suppose that for some $t \in [0,1]$, $f(t) \notin A'$, from which it follows that $f(t) \in A$. Then by Proposition A.1 and the definition of F,

$$|x' - x|_2 = |f(0) - f(t)|_2 + |f(t) - f(1)|_2 \geq \frac{2\ell b}{\lambda}.$$

Then we have

$$|u(x) - u(x')| \leq |u(x)| + |u(x')| \leq 2\ell b \leq \lambda |x - x'|_2,$$

and (A.30) holds again. This completes the proof.

Appendix B

Additional Material Relating to Chapter 5

In several papers using the concept of approximately finite memory (for example, [26] and [53]), reference is made in the introductions to two "related" senses in which a system could be said to have approximately finite memory. One sense is used in the body of those papers, and in this chapter. The other sense appears in [11], and is closely related to the early definition used in [10]. We refer to this sense below as property $\mathcal{A}_+$.

In fact, approximately finite memory (as defined in this chapter) is equivalent to property $\mathcal{A}_+$. This is neither stated nor shown in earlier papers. Further, because it is often more convenient to show that a map has property $\mathcal{A}_+$, property $\mathcal{A}_+$ may serve as a useful condition to show that a system has approximately finite memory.

In this appendix, we use the notation of Chapter 5, as well as the following. We say that $G : S_+ \to V_+$ has property $\mathcal{A}_+$ if for every $\varepsilon > 0$, there exists a $\Delta \geq 0$ such that

$$|(Gu)(t) - (GW_{t,\Delta}u)(t)| < \varepsilon$$

for every $t \in \mathbb{R}_+$ and for every $u \in S_+$.

Proposition B.1: $G : S_+ \to V_+$ has property $\mathcal{A}_+$ if and only if G has approximately finite memory.

Proof: If G has approximately finite memory, it is immediately obvious that it has property $\mathcal{A}_+$. Conversely, suppose that G has property $\mathcal{A}_+$. Let $\varepsilon > 0$, and let Δ be the associate of $\frac{1}{2}\varepsilon$ in property $\mathcal{A}_+$. Let $u \in S_+$, $t \in \mathbb{R}_+$, and $\alpha > \Delta$. By the triangle inequality,

$$|(Gu)(t) - (GW_{t,\alpha}u)(t)| \leq |(Gu)(t) - (GW_{t,\Delta}u)(t)| + |(GW_{t,\Delta}u)(t) - (GW_{t,\alpha}u)(t)|.$$

Using property $\mathcal{A}_+$, each term on the right side above is bounded by $\frac{1}{2}\varepsilon$. Therefore,

$$|(Gu)(t) - (GW_{t,\alpha}u)(t)| < \varepsilon,$$

showing that G has approximately finite memory. This completes the proof.

Bibliography

[1] Bill Toole, Calvin Plett, and Mark Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETs," *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 51, no. 2, pp. 319–328, Feb. 2004.

[2] Domenico Mirri et. al., "A nonlinear dynamic model for performance analysis of large-signal amplifiers in communications systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 2, pp. 341–350, 2004.

[3] Piergiovanni Marzocca, Walter A. Silva, and Liviu Librescu, "Nonlinear open-/closed-loop aeroelastic analysis of airfoils via Volterra series," *AIAA Journal*, vol. 42, no. 4, pp. 673–686, Apr. 2004.

[4] Richard Tymerski, "Volterra series modeling of power conversion systems," *IEEE Transactions on Power Electronics*, vol. 6, no. 4, pp. 712–718, 1991.

[5] Constantine Kotropoulos and Ioannis Pitas, Eds., *Nonlinear Model-Based Image/Video Processing and Analysis*, John Wiley and Sons, New York, NY, 2001.

[6] Stephen Boyd and Leon O. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Transactions on Circuits and Systems*, vol. 32, no. 11, pp. 1150–1161, Nov. 1985.

[7] Irwin W. Sandberg, "A perspective on system theory," *IEEE Transactions on Circuits and Systems*, vol. 31, no. 1, pp. 88–103, 1984.

[8] F. Fiori and P. S. Crovetti, "Prediction of EMI effects in operational amplifiers by a two-input Volterra series model," *IEE Proceedings: Circuits, Devices, and Systems*, vol. 150, no. 3, pp. 185–193, June 2003.

[9] Martin Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley, New York, 1980.

[10] Irwin W. Sandberg, "Structure theorems for nonlinear systems," *Multidimensional Systems and Signal Processing*, vol. 2, no. 3, pp. 267–286, 1991.

[11] Irwin W. Sandberg, "Uniform approximation and the circle criterion," *IEEE Transactions on Automatic Control*, vol. 38, no. 10, pp. 1450–1458, Oct. 1993.

[12] Irwin W. Sandberg and Lilian Xu, "Approximation of myopic systems whose inputs need not be continuous," *Multidimensional Systems and Signal Processing*, vol. 9, no. 2, pp. 207–225, Apr. 1998.

[13] Irwin W. Sandberg, James T. Lo, Craig L. Fancourt, Jose C. Principe, Shigeru Katagiri, and Simon Haykin, *Nonlinear Dynamical Systems: Feedforward Neural Network Perspectives*, Wiley-Interscience, New York, NY, 2001.

[14] Micheal B. Matthews, "Approximating nonlinear fading-memory operators using neural network models," *Circuits, Systems, and Signal Processing*, vol. 12, no. 2, pp. 279–307, June 1993.

[15] D. J. Newman and H. S. Shapiro, "Jackson's theorem in higher dimensions," in *On Approximation Theory: Proceedings of the Conference held in the Mathematical Research Institute at Oberwolfach, Black Forest, August 4–10, 1963*, P.L. Butzer and J. Korevaar, Eds., Germany, 1964, pp. 208–217, Birkhäuser Verlag Basel und Stuttgart.

[16] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.

[17] Jooyoung Park and Irwin W. Sandberg, "Approximation and radial basis function networks," *Neural Computation*, vol. 5, no. 2, pp. 305–316, 1993.

[18] Ward Cheney and Will Light, *A Course in Approximation Theory*, Brooks/Cole Publishing Company, Pacific Grove, CA, 2000.

[19] M. H. Stone, "A generalized Weierstrass approximation theorem," in *Studies in Modern Analysis*, vol. 1, pp. 30–87. Prentice-Hall, Englewood Cliffs, NJ, 1962, R. C. Buck, editor.

[20] Irwin W. Sandberg, "Approximation theorems for discrete-time systems," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 564–566, May 1991.

[21] Micheal Ciraula and Irwin W. Sandberg, "Uniform approximation of discrete-time nonlinear systems," in *Proceedings of the 1999 International Joint Conference on Neural Networks*, Washington, D.C., July 1999, On CD.

[22] Irwin W. Sandberg, "Time-delay neural networks, Volterra series, and rates of approximation," *Circuits, Systems, and Signal Processing*, vol. 17, no. 5, pp. 653–665, 1998.

[23] Irwin W. Sandberg, "Constructive approximation of non-linear discrete-time systems," *International Journal of Circuit Theory and Applications*, vol. 28, pp. 109–120, 2000.

[24] P. J. Broser et. al., "Nonlinear anisotropic diffusion filtering of three-dimensional image data from 2-photon microscopy," in *Proceedings of SPIE-IS and T Electronic Imaging — Image Processing: Algorithms and Systems IV*,

vol. 5672 of *Proceedings of the SPIE — The International Society for Optical Engineering*, pp. 44–69. SPIE, Bellingham, WA, 2005.

[25] Takahiro Saito, Jun Satsumabayashi, and Takashi Komatsu, "Selective image-sharpness enhancement by coupled nonlinear reaction-diffusion time-evolution equations and its application to suppression of breathing distortions in moving pictures," *Electronics and Communications in Japan Part III: Fundamental Electronic Science (English Translation of Denshi Tsushin Gakkai Ronbunshi)*, vol. 87, no. 5, pp. 33–47, 2004.

[26] Irwin W. Sandberg, "Approximately finite memory and input-output maps," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 39, no. 7, pp. 549–556, 1992.

[27] Irwin W. Sandberg, "Uniform approximation of time-invariant non-linear systems," *International Journal of Circuit Theory and Applications*, vol. 32, pp. 105–116, 2004.

[28] Richard L. Wheeden and Antoni Zygmund, *Measure and Integral*, Marcel Dekker, New York, NY, 1977.

[29] L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis*, Frederick Ungar Publishing Company, New York, NY, 1961.

[30] George Bachman, Lawrence Narici, and Edward Beckenstein, *Fourier and Wavelet Analysis*, Springer-Verlag, New York, 2000.

[31] Gabor Szegö, *Orthogonal Polynomials*, vol. XXIII of *American Mathematical Society Colloquium Publications*, American Mathematical Society, New York, NY, 1939.

[32] Bateman Manuscript Project, *Higher Trancendental Functions*, vol. 2, McGraw-Hill Book Company, New York, NY, 1953.

[33] Irwin W. Sandberg, "Some results on the theory of physical systems governed by nonlinear functional equations," *Bell System Technical Journal*, vol. 44, pp. 871–898, May 1965.

[34] Irwin W. Sandberg, "On the L_2-boundedness of solutions of nonlinear functional equations," *Bell System Technical Journal*, vol. 43, pp. 1581–1599, July 1964.

[35] Federico Girosi, "Approximation error bounds that use VC-bounds," in *Proceedings of the International Conference on Artificial Neural Networks*, Paris, France, 1995, pp. 295–302, also available at http://www.svms.org/vc-dimension/Giro95.pdf.

[36] Andrew R. Barron, "Neural net approximation," in *Proceedings of the 7^{th} Yale Workshop on Adaptive and Learning Systems*, K. S. Narendra, Ed., May 1992, pp. 69–72, also available at http://www.stat.yale.edu/~arb4/publications_files /neural%20net%20approximation.pdf.

[37] Lee K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *The Annals of Statistics*, vol. 20, no. 1, pp. 608–613, 1992.

[38] Andrew R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, May 1993.

[39] Sanjit K. Mitra and Giovanni L. Sicuranza, Eds., *Nonlinear Image Processing*, Academic Press, San Diego, CA, 2001.

[40] Petros Maragos, "Lattice image processing: A unification of morphological and fuzzy algebraic systems," *Journal of Mathematical Imaging and Vision*, vol. 22, no. 2-3, pp. 333–353, 2005.

[41] Jyh-Yeong Chang and Jia-Lin Chen, “Classifier-augmented median filters for image restoration,” *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 2, pp. 351–356, 2004.

[42] G. Ramponi, S. Marsi, and S. Carrato, “Image processing using rational functions,” in *Nonlinear model-based image/video processing and analysis*, chapter 3, pp. 59–91. John Wiley and Sons, New York, NY, 2001.

[43] Irwin W. Sandberg and Lilian Xu, “Uniform approximation of discrete-space multidimensional myopic maps,” *Circuits, Systems, and Signal Processing*, vol. 16, no. 3, pp. 387–403, 1997.

[44] Irwin W. Sandberg and Lilian Xu, “Uniform approximation of multidimensional myopic maps,” *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 44, no. 6, pp. 477–485, June 1997.

[45] Luis Miguel Sanchez-Brea and Eusebio Bernabeu, “Determination of the optimum sampling frequency of noisy images by spatial statistics,” *Applied Optics*, vol. 44, no. 16, pp. 3276–3283, 2005.

[46] Tom M. Apostol, *Mathematical Analysis*, Addison-Wesley, Reading, MA, 1974, second edition.

[47] Eva Matoušková, “Extensions of continous and lipschitz functions,” *Canadian Mathematical Bulletin*, vol. 43, no. 2, pp. 208–217, 2000.

[48] Nelson Dunford and Jacob T. Schwartz, *Linear Operators, Part I: General Theory*, Interscience Publishers, New York, NY, 1958.

[49] Andreas Poncet, Jean L. Poncet, and George S. Moschytz, “On the input-output approximation of nonlinear systems,” in *Proceedings – IEEE International Symposium on Circuits and Systems*, Seattle, WA, 1995, vol. 2, pp. 1500–1503.

[50] Jeff S. Shamma and Rongze Zhao, "Fading-memory feedback systems and robust stability," *Automatica*, vol. 29, no. 1, pp. 191–200, 1993.

[51] Munther A. Dahleh, Eduardo D. Sontag, David N. C. Tse, and John N. Tsitsiklis, "Worst-case identification of nonlinear fading memory systems," *Automatica*, vol. 31, no. 3, pp. 503–508, 1995.

[52] Irwin W. Sandberg, "The mathematical foundations of associated expansions for mildly nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 7, pp. 441–454, July 1983.

[53] Irwin W. Sandberg, "Approximately finite memory and the circle criterion," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 41, no. 7, pp. 473–476, July 1994.

[54] Irwin W. Sandberg, "$\mathbb{R}_+$ fading memory and extensions of input-output maps," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 49, no. 11, pp. 1586–1591, Nov. 2002.

[55] Lawrence C. Evans and Ronald F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.

[56] Erwin Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley and Sons, New York, NY, 1978.

Vita

Mark Allan Story was born in Metairie, Louisiana on May 18, 1973, the son of Francis Allan Story and Linda Kay Story. He graduated from Cumberland Regional High School, Seabrook, New Jersey in 1991 and entered the Massachusetts Institute of Technology. During the summers of 1993, 1994, and 1995, and the fall of 1995, he participated in the 6A internship program, with employment at Tektronix in Beaverton, Oregon. He was awarded a Bachelor of Science degree from the Massachusetts Institute of Technology in June 1996, and a Master of Engineering degree from the Massachusetts Institute of Technology in June 1997. His masters thesis was entitled "Multiplierless Decimation and Commercial Postfiltering of a Discrete-Time Signal." He spent the following summer employed at Rockwell-Collins in Cedar Rapids, Iowa. In September 1997, he entered the Graduate School of The University of Texas at Austin, receiving a Microelectronics and Computer Development Fellowship and a Temple MCD Fellowship. He is currently employed full-time as a Engineering Scientist Associate at Applied Research Laboratories, Austin, Texas.

Permanent Address: 4111 Kilgore Lane
Austin, TX 78727

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.